THESIS

OBJECT AND ACTION DETECTION METHODS USING MOSSE FILTERS

Submitted by

Robert T. Arn

Department of Mathematics

In partial fulfillment of the requirements

For the degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2012

Master's Committee:

Advisor: Michael Kirby
Co-Advisor: Chris Peterson

Bruce Draper

ABSTRACT

OBJECT AND ACTION DETECTION METHODS USING MOSSE FILTERS

In this thesis we explore the application of the Minimum Output Sum of Squared Error (MOSSE) filter to object detection in images as well as action detection in video. We exploit the properties of the Fourier transform for computing correlations in two and three dimensions. We perform a comprehensive examination of the shape parameters of the desired target response and determine values to optimize the filter performance for specific objects and actions. In addition, we propose the Gaussian Iterative Response (GIR) algorithm and the Multi-Sigma Geometric Mean method to improve the MOSSE filter response on test signals. Also, new detection criteria are investigated and shown to boost the detection accuracy on two well-known data sets.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Automation of information extraction from persistence surveillance data in images and video has numerous applications for defense, medical care, transportation, and many other disciplines. Object and action recognition technology would allow for continuous monitoring of national borders, heavy crime areas, and public areas on an unprecedented scale. Instead of requiring a large number of people each watching multiple video feeds in order to be able to detect anomalous objects or actions, we could instead have computer systems analyzing the data feeds in real time. This thesis addresses one of the main challenges related to developing a computer surveillance system, i.e. automatic object detection in images and action detection in video.

We can use image processing techniques to select and modify parts of signals for the goal of object detection. Every image taken by digital cameras has the potential to contain numerous objects such as trees, boats, or people. By examining these images, it is fairly easy for the human brain to recognize the representation of these objects in a digital image. However, describing what the representation of a generalized object is to a computer is a difficult task. Consider the situation where an ATM machine takes an image every few seconds. It would be advantageous to be able to determine if, e.g., there is a gun within the field of view. If so, the authorities need to be notified immediately. Unless there is a way to automatically detect the shape and size of a gun through a series of computer algorithms, the only way for this detection to occur is if a human manually looks at each image taken from every ATM machine. By creating a model of the appearance of a gun in terms that a computer can understand, it would be possible to automatically search through every ATM image for that model and flagging an alert when the object is detected. Instead of requiring hundreds of thousands of man-hours per month to view the images, the image evaluation process could be done automatically using computers.

Another problem central to automated surveillance system is action detection in video. It would be advantageous to be able to detect particular actions taken by humans (or other subjects) whom have been recorded onto digital video. By generating a model of what the action 'throw' appears when projected onto a video, we can quickly search a set of videos for this action. Given that a large number of building, streets, and homes have security cameras that currently only serve to review known crimes, we could instead use this technology to alert us, in real time, when a potential crime is occurring by detecting actions such as raising a weapon, one person striking another, or kicking down a door.

To begin to understand signal detection, we examine a template matching approach. Template matching refers to attempting to detect a feature given as a model in a test signal. The simplest detection approach is to start with a model, n, which contains our desired feature and a test signal t which may or may not contain the feature we are trying to detect. By vectorizing these signals, we then compute the dot product:

$$v = t \cdot n.$$

If $v > \epsilon$, where ϵ is a threshold value, then a detection is said to occur. One problem with this very basic approach to object detection is the location of the feature in the model must essentially match the location of the feature in the test image to obtain optimal results.

To help solve this problem, we start looking at correlation based template matching filters. In general, these filters operate on the same principle. A model of the desired feature is created, then the model is applied to the test image. But instead of taking a dot product between t and n we correlate them, meaning that we search t for the best representation of n through a series of horizontal, vertical, and rotational shifts. The resulting output

$$v = t \otimes n$$

is described by each pixel of v representing how well the region surrounding the corresponding pixel of t is similar to the model, n, where $\otimes$ is the correlation operator. This removes the condition that the pixel location of the feature in the model must match the pixel location of the feature in the test image. One of the remaining problems is how do we create a model of the feature we are trying to detect?

MOSSE filters provide us with a method of creating a model from real data [2, 3, 4]. By identifying a target feature in a handful of images or videos, we generate a MOSSE filter, m^*. This filter strongly weights the frequencies local to the identified point while weakly averaging the rest of the signal. By generating the MOSSE filter from real data, we can obtain more accurate results from searching not only for the target feature, but characteristics of the feature such as proximity to other fixed features, feature translations, and background information. As we describe in Chapter 3, MOSSE filters extend naturally to higher dimensions. We will show the use of these filters for the problems of object detection (two-dimensional filters) and action detection (three-dimensional filters).

In two-dimensions, MOSSE filters are used to explore the problem of object detection. Given a series of digital images (with height and width), we will spend Chapter 4 creating four MOSSE filters, each designed to detect a particular facial feature in a series of images all containing a human face. We will explore the creation of the MOSSE filter in detail and show how to optimize the creation of the filter by solving the minimization problem which leads to the creation of m^*. Given that m^* depends on a number of parameters, such as the size and shape of a Gaussian filter, we will examine the parameter space to determine the optimal Gaussian filter to use for each facial feature. Once the MOSSE filter is applied to a series of test images, we will explore the properties of the response through a series of new techniques such as Gaussian Iterative Responses, Multi-Sigma Geometric Mean, Peak-to-Side-Lobe Ratios, and maximized pixel intensities.

Finally, in Chapter 5 we will show the extension of MOSSE filters into three-dimensions. Working with the KTH data set, a series of videos (dimensions of height, width, and time)

each containing a human performing an action, we will explore the difficulties associated with action detection. In particular, we will show attempts to identify the optimal parameters for the creation of the Gaussian filter to use in the creation of the MOSSE filter. After 3-D MOSSE filters have been created for all 6 actions contained in the data set, we will show the responses obtained by applying the filter to test videos and we will examine the issue of minimizing false detections. Concluding our current action detection attempts, we will outline a series of issues to overcome in order to obtain a high accuracy detection rate.

Chapter 2

DESCRIPTION OF THE DATA SETS

Two well known computer vision data sets were used for this project: FERET and KTH. These data sets showcase the power of creating and applying the MOSSE filter on real world data.

2.1 FERET

The Facial Recognition Technology (FERET) data set is comprised of 14051 still images release by the Defense Advanced Research Products Agency (DARPA) through the Department of Defense's Counterdrug Technology Development Program [12]. Each image is 384 pixels tall by 256 pixels wide, is presented in 8-bit grayscale, and contains a human face at various angles to the camera. These faces are presented under different lighting conditions and with variations in facial hair, race, gender, and the presence of glasses. Many images contain partially open or closed eyes. Ground truth data is provided by DARPA where the center of each eye, the center of the nose and the center of the mouth are given. The full data set is broken down into parts by release date. For this project we used a subset of the data, taking only 3364 images into account.

Each image has been preprocessed (scaled, rotated, and cropped) from the original camera data so the eyes are fairly evenly spaced and the eyes, center of nose, and center of the lips are in roughly the same pixel location. For reference, given that row 1, column 1 is the upper-left corner of an image, the right eye is located around row 180, column 100. The left eye, nose, and mouth are located around rows 180, 220, and 250, columns 180, 140, and 140 respectively.

For future reference, we have computed the average distance between the 4 facial features using the l_∞-norms in Table 2.1.

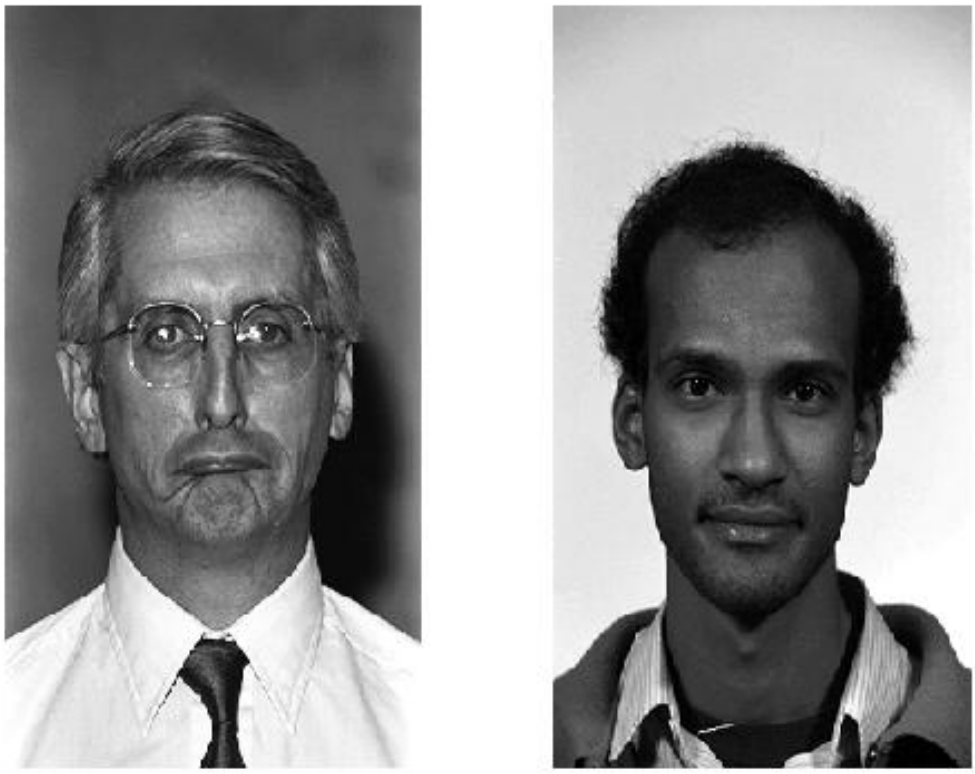

Figure 2.1: Two examples FERET raw images.

2.2 KTH

The KTH data set is comprised of 599 videos, each 120 pixels tall by 160 pixels wide by 256 frames. In this data set, there are 25 human subjects, performing one action per video (boxing, handwaving, handclapping, jogging, running, or walking) as seen in Figure 2.2, under four conditions (outdoors, outdoors with scale variations, outdoors with different clothes, and indoors with harsh lightning) as seen in Figure 2.3. In order to make this data set easier to work with, we duplicated one of the handclapping videos to replaced the corrupted video from the original data. This action gave us a 600 video data set, with 100 videos per action.

Table 2.1: l_∞-norm distances between target facial features.

	R. Eye	L. Eye	Nose	Mouth
R. Eye	0	67	33	71
L. Eye	67	0	34	71
Nose	33	34	0	43
Mouth	71	71	43	0

No ground truth data has been provided by the authors of this data set. To effectively use the algorithms presented in this thesis, we have made ground truth data. In order to reproduce results, our ground truth data is given in Appendix A.

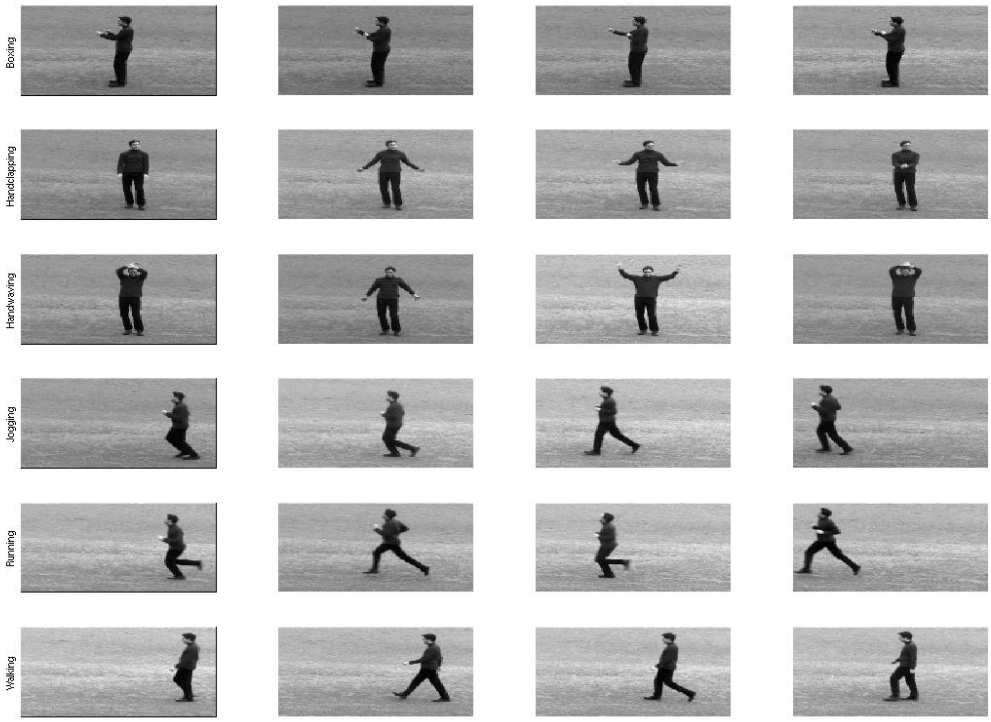

Figure 2.2: KTH Action Classes.

Figure 2.3: KTH Scenarios.

Chapter 3

MOSSE FILTERS

3.1 Pre-Processing the Data

MOSSE filters create a model of the target feature by using a set of training data. However, before we can use this data in the construction of a filter we need to apply some pre-processing techniques in order to normalize the data. We start by defining the set f to be the elements $f_1, f_2, f_3...$ where each f_i is a training signal, or a signal to be used in the creation of the filter. Now we perform the following operations:

1. Mean subtract the data. For each f_i define m_i to be the mean across all pixels of f_i then take

$$\overline{f}_i = f_i^p - m_i$$

where f_i^p is the p^{th} pixel index of f_i.

2. Set the standard deviation of the mean subtracted data, now called f_i, to 1 by finding h_i, the global maximum of f_i and computing

$$\overline{f}_i = \frac{f_i^p}{h_i}.$$

3. Finally, we will apply a cosine window to the mean subtracted signal with a standard deviation of 1, now called f_i. Defining n to be the dimension of a single signal, we construct a n-dimensional discretized cosine window of one period by

$$c(i,j) = \sin\left(\frac{2\pi i}{L_1}\right)\sin\left(\frac{2\pi j}{L_2}\right)$$

or

$$c(i,j,k) = \sin\left(\frac{2\pi i}{L_1}\right)\sin\left(\frac{2\pi j}{L_2}\right)\sin\left(\frac{2\pi k}{L_3}\right)$$

depending on the dimension of the data. Here, L_1 is the number of rows which i is an index over, L_2 is the number of columns which j is an index over, and L_3 is the number of frames which k is an index over. where j_k is the j^{th} pixel along the k^{th} dimension and x_k is the length of the data along the k^{th} dimension. Finally, we perform a point-wise multiplication between the cosine window and the mean subtracted, standard deviation adjusted data.

These operations are designed to normalize the data to help reduce changes in lighting conditions while minimizing potential artifacts created by the geometry of the MOSSE filter. As we will discuss later in Chapter 3.3, we will only apply this series of preprocessing to the training data, not the testing data.

3.2 Constructing a MOSSE Filter

Define the set f as above. Then define the set s_i associated with each $f_i \in f$ to be the set of target points for f_i.

Our mask filter w_i is the desired response in physical space and is given by

$$w_i(x) = \sum_{\{j|x_j \in S_i\}} g_j(x)$$

where

$$g_j(x) = \frac{1}{2\pi\sigma^{n/2}} \exp\frac{-||x-x_j||_2^2}{2\sigma^2}$$

and n is the dimension of the data.

There are 3 possibilities for w_i. The first is that $w_i = 0$ which occurs when we have no target points associated with an f_i. This will allow us to train the filter to not respond

to certain frequencies corresponding to targets which we do not desire to detect. Another possibility is that there only exists one target point for a given f_i, in which case $w_i = g$. The last case is when there are multiple target points for a given f_i. In this case, w_i will be a sum of Gaussians where each Gaussian is centered at a unique target point.

To increase computational efficiency we will perform the remaining calculations in frequency space. To start, we will denote:

$$F_i = \mathcal{F}(f_i)$$

and

$$W_i = \mathcal{F}(w_i).$$

Here we take note that the Fourier Transform of a Gaussian is a Gaussian - a property that this filter utilizes heavily as we move between pixel and Fourier space. To further decrease computation time, we will use a Fast Fourier Transformation [13].

The goal of these computations is to obtain our desired response signal, W_i by requiring

$$W_i = F_i \odot M^*$$

where M is an unknown filter and $\odot$ is component-wise multiplication. In pixel space the problem we seek to solve is

$$\underset{m^*}{\text{minimize}} \sum_i ||w_i - f_i \otimes m^*||^2.$$

The frequency space representation of this optimization problem is

$$\underset{M^*}{\text{minimize}} \sum_i ||W_i - F_i \odot M^*||^2.$$

An equation for M^* may be obtained by setting the derivative of the objective with respect to M^* to zero, i.e. [10]:

$$\frac{\partial}{\partial M}\sum_i (W_i - F_i \odot M, W_i - F_i \odot M) = 0$$

$$\frac{\partial}{\partial M}\sum_i (W_i^* \odot W_i + (F_i \odot M)^* \odot F_i \odot M - W_i^* \odot F_i \odot M - (F_i \odot M)^* \odot W_i) = 0$$

$$\frac{\partial}{\partial M}\sum_i (W_i^* \odot W_i + M^* \odot F_i^* \odot F_i \odot M - W_i^* \odot F_i \odot M - (M^* \odot F_i^*) \odot W_i) = 0$$

$$\sum_i (M^* \odot F_i^* \odot F_i - W_i^* \odot F_i) = 0$$

$$\sum_i (M^* \odot F_i^* \odot F_i) - \sum_i (W_i^* \odot F_i) = 0$$

$$\sum_i (W_i^* \odot F_i) = M \odot \left(\sum_i (F_i^* \odot F_i)\right).$$

So in solving for our MOSSE filter,

$$M = \frac{\sum_i (W_i^* \odot F_i)}{\sum_i (F_i^* \odot F_i)}$$

or

$$M^* = \frac{\sum_i (W_i \odot F_i^*)}{\sum_i (F_i \odot F_i^*)}$$

where division is point-wise division. The reason for using M^* versus M is computational. In M^* we do not need to compute the conjugate of the filter W_i.

3.3 Applying MOSSE

Once we have our MOSSE filter, M^*, we can introduce a new test signal, t, and apply the filter to it in an attempt to get a response which is similar in structure to g. Given a single target point per n-dimensional data, the MOSSE filter, when mapped back into pixel

space by

$$m^* = \mathcal{F}^{-1}\left(M^*\right)$$

assuming the target locations used represent a similar feature across the entire training data set, yields an outline of the desired featured centered in the n-dimensional data structure. Given multiple target points per training data, m^* has an average of all target features centered in the data structure. Figure 3.1 shows us examples of MOSSE filters when trained on various facial features from the FERET data set. Notice the target feature is strong, yet the structure of the rest of the face is still present.

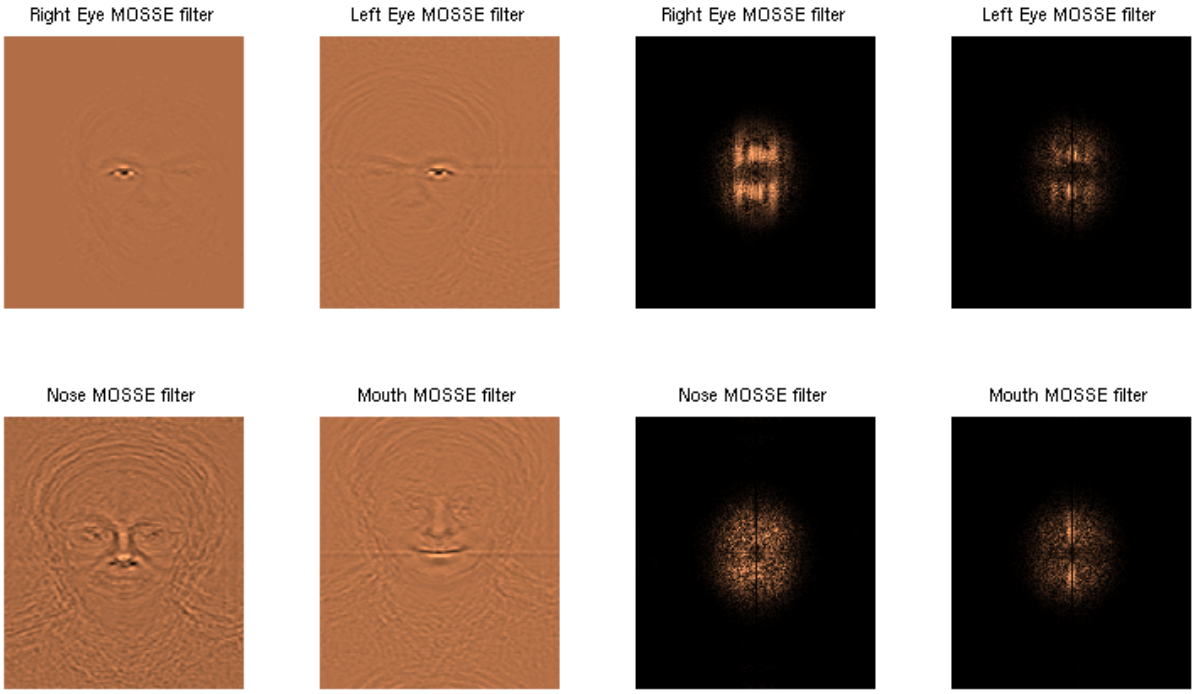

Figure 3.1: MOSSE Filters defined on various targets in FERET data.

Preprocessing of the training data, with the cosine window, removes unwanted high frequency information. However, we observe that when we apply the cosine window to testing data, the response signal is much weaker. This is especially true when the desired action/object is near the boundary of the data. For example, consider a data cube associated with a video of an action. Application of the cosine filter to this effectively removes information at the edges of the cube. Hence, we have opted to not filter the test data and

risk failing to detect actions near the boundary of the data cube. Similar considerations hold for 2-D images. Thus we limit our preprocessing to mean subtraction and normalizing to unit standard deviation.

To apply the MOSSE filter (in Fourier space) we compute a pointwise multiplication of T, the testing signal in frequency space, and the MOSSE filter, M^*, thereby getting a response signal

$$R_1 = T \odot M^*$$

.

By computing r_1, the inverse Fourier Transform of R_1, we can visually locate regions of the new signal t which are similar to the trained targets.

Chapter 4

OBJECT DETECTION

Given a MOSSE filter that has been trained on a target feature, our task is to apply the filter to a set of test data and identify if the desired target feature is contained within the image, and if so, where the target feature is located. In this chapter, we will explore the creation of a 2-D MOSSE filter on the FERET data set. Techniques for optimizing the performance of the creation of the filter as well as techniques designed to operate on the response signal generated by applying a MOSSE filter to a test image will be explored. In addition, we will present a series of detection criteria and show detection results with these methods.

4.1 Measuring Squared Error

To create the MOSSE filter in Chapter 3 we solved the minimization problem

$$\underset{M^*}{\text{minimize}} \sum_i ||W_i - F_i \odot M||^2.$$

We note that using filters for object detection that minimize the sum of squared errors between the output of the application of the filters and the training filter is not a new idea [6, 9]. Here however, we will explore this idea in the context of MOSSE filters as we vary parameters used in its construction. In particular, we desire to obtain a connection between the construction of a filter, the size, shape, and representation of the target feature, and the numerical value for the sum of squared error.

For the case of object detection, we start by working with the FERET data set as described in Chapter 2.1. For the rest of this chapter, unless otherwise stated, we will split this data set into 2 parts, the first 512 images of the data set will be used to create the MOSSE filter, here after called the training data, and the remaining 2852 images, here after

called the testing data, will be used to test the filter. We chose to use 512 images for training purposes based upon prior filter experiments [1]. The entire data set has been preprocessed as described in Chapter 3.1.

Next, since σ is our only parameter for the MOSSE filter, we would like to determine the effects of varying σ. Recall that we create a mask filter

$$w_i(x) = \sum_{j|x_j \in S_i} g_j(x)$$

where

$$g_j(x) = \frac{1}{2\pi\sigma^{n/2}} \exp^{\frac{-||x-x_j||_2^2}{2\sigma^2}}.$$

We choose a step size for σ of 0.1, and create a MOSSE filter using the training data set with unique σ value from 1 to 10.9. Then we apply the MOSSE filter over the 2852 images in the training data set.

In Figure 4.1 we show the results of

$$\underset{M^*}{\text{minimize}} \sum_i ||W_i - F_i \odot M^*||^2.$$

computed in several different ways.

The top-left plot shows the results of the above computations, where the sum squared error is computed over the entire pixel response, r_1. The sum squared error for both of the eyes show clear structure i.e. when we select a value of σ around 2.8 the error is minimized. However, the nose and mouth do not show a similar structure. A hypothesis as to why can be found in Chapter 4.2.

The top-right plot computes the sum squared-error when measured in an 11x11 pixel window around the given ground truth information. We note that as σ increases the error decreases. This result is to be expected as the majority of the error is created from the sum of false-positive detections. We show this is the case as we examine the average maximum

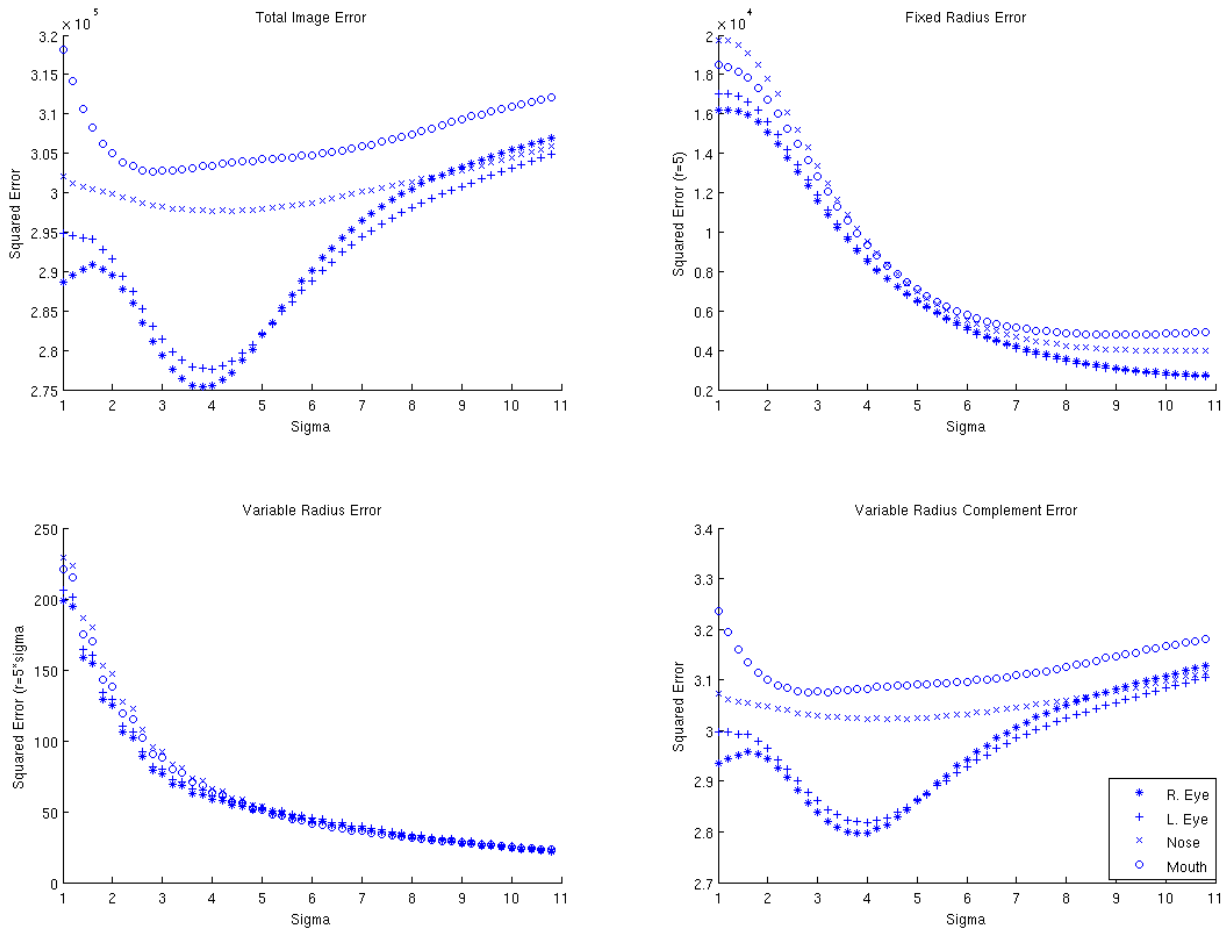

Figure 4.1: FERET data sum of squared error.

pixel intensity value in Figure 4.2. When the size of the Gaussian increases, the intensity values of the pixel response decrease as shown in Figure 4.1.

Examining the bottom-left plot shows similar results. This time, we varied the radius around the ground truth data so in all cases, we were examining a 5σ x 5σ window around the given ground truth. In order to compare the results from σ value to σ value, we then divided the sum by the number of pixels in the window. Note once again, the error decreases monotonically as the Gaussian size increases.

Finally, the bottom-right plot of Figure 4.1 shows the complement of the bottom-left plot. Notice, the structure is very similar to the error calculated over the entire image. Note that the average maximum intensity pixel value, the fixed radius summed squared error, and the variable radius averaged squared error (when properly scaled back to a sum instead of

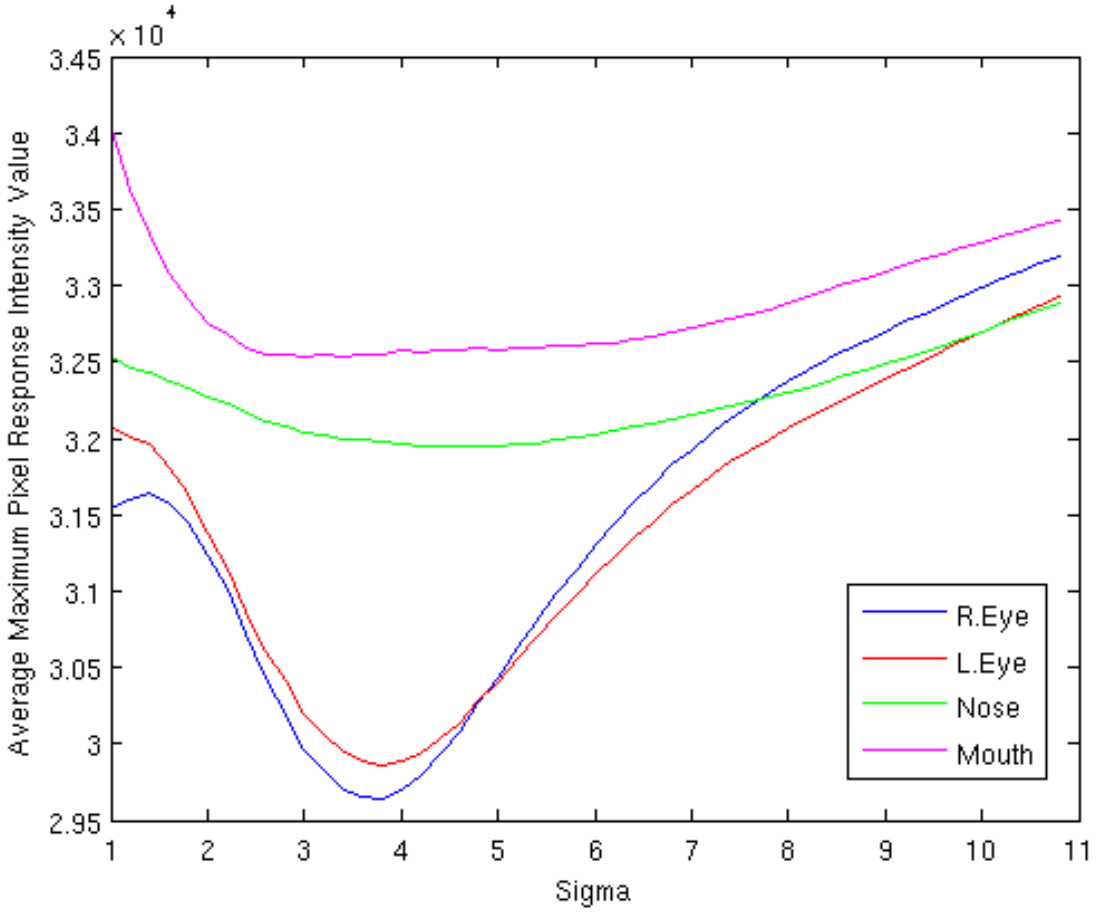

Figure 4.2: FERET average maximum intensity pixel value.

an average) are all an order of magnitude less than the total image error and the variable complement error (again, once scaled back to a sum instead of an average). This leads us to conclude that the majority of the squared error between the filter and the desired (trained) response comes from false-positives or places in the testing data where the filter responds to an non-target feature. By using a σ where this summed squared error is minimized, we hope to minimize the amount and intensity of misfires by the MOSSE filter onto the testing data, thereby improving the filter response.

4.2 Selecting Sigma

The goal of creating and applying a MOSSE filter to a test data set is to be able to identify the target feature by examining the response r_1. We expect there there will be a

strong response in the pixel location around the identified target and a weak response in all pixel locations not around the identified region. However, the result in real-world (non-synthetic) examples is not as clear. As we see in Figure 4.3, the pixel response generated by using a MOSSE filter created using a spherical Gaussian for g_j with $\sigma = 2$ centered on the right eye and applied to a test image of a face containing a right eye of roughly the same pixel size, we not only get a strong response around the area corresponding to the right eye in the test data, but we also get a non-zero response on the other (left) eye, and other parts of the face/body.

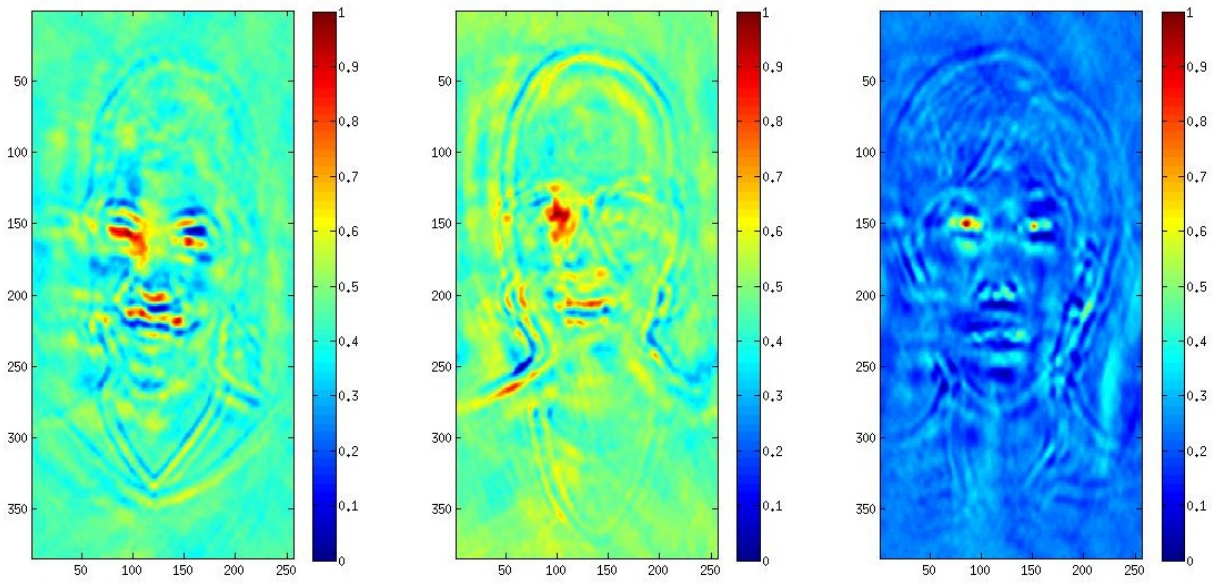

Figure 4.3: Pixel responses with right eye trained MOSSE filter.

As we concluded from Section 4.1, on the well-defined eye features we can minimize the false detections by carefully selecting a value for σ in the creation of g_j. The idea behind this is, in object detection, the size of the target object and surrounding detail helps to determine the optimal σ that should be used. For example, if we have an eye detector and the size of an eye is known to have a diameter of 20 pixels, then we will want to select a σ which generates a Gaussian large enough to strongly weight the entire eye, yet not so large as to strongly weight the nose (or ear, or other image feature).

At this point, we take note that there is not a strong global minimum for the sum of squared error on the nose. Examining the ground truth data for the nose may yield some possible explanations. When compared with the eyes, the nose is a poorly defined feature on a face. In the FERET data, the ground truth on the nose is defined to be the center of tip on the nose. However, this low contrast feature can be difficult to detect (by humans creating the ground truth data). In addition, due to the pre-scaling of the images, the eyes are roughly the same size. In general, the size and shape of the nose has a greater variation. Since the MOSSE filters are not scale invariant, if the MOSSE filter is trained on noses of various sizes and applied to data of various sizes, we expect there will be weaker intensity values for positive detections and stronger intensity values for some false-positive detections. While we do see some structure within the summed squared error for the mouth, the same argument applies.

To help solidify our argument, we determine the number of successes on the test data set. Here, we consider the detection a failure if the maximum intensity value of

$$r_1 = \mathcal{F}^{-1}\left(T \odot M^*\right)$$

is located within 20 pixels of the ground truth data using the l_∞-norm; otherwise it is considered to be a failure. We chose to use a radius of 20 pixels as that was the radius used in [1]. This value will allow us to avoid all but the most extreme (if any) examples of mislabeling of ground truth data.

In Figure 4.4 we give the total number of failures when the MOSSE filter, made with Gaussians of various σ values, is applied to both the training data set (of 512 images) and the testing data set (of 2852 images). The results of both of these are presented to show there are minimal differences in final results when the filter is applied to the data it was trained on, provided the training size is large enough.

To perform this experiment, we trained 100 MOSSE filters, M^*, on the same training data set. For each MOSSE filter we varied the value of σ in

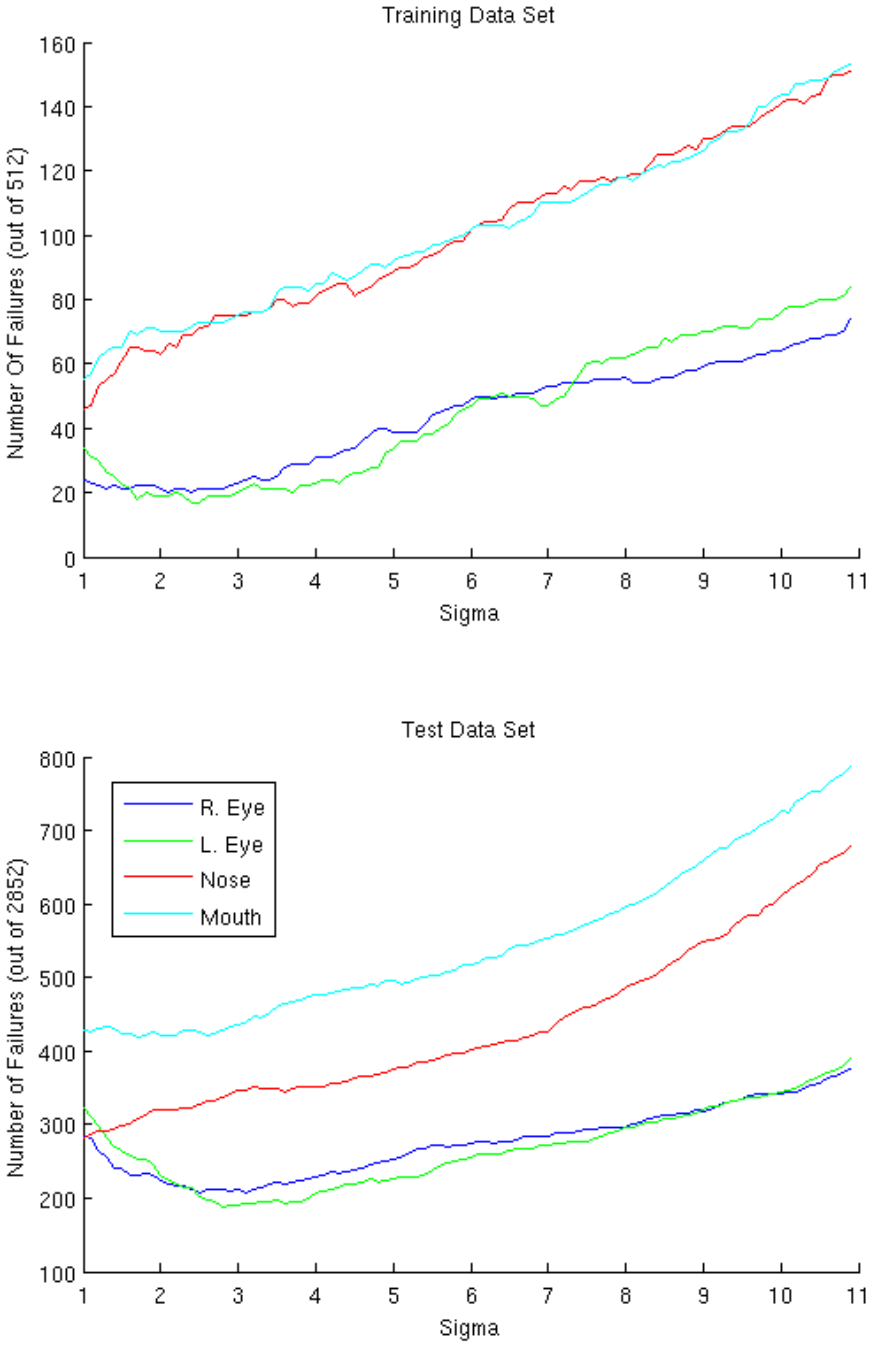

Figure 4.4: FERET failures.

$$g_j(x) = \exp^{\frac{-||x-x_j||_2^2}{2\sigma^2}}$$

by $\sigma_{\text{step}} = 0.1$. This step size allows enough variability in the Gaussian filter to make non-trivial numerical differences in the created MOSSE filters.

A summary of the optimal value for σ is found in 4.1.

Table 4.1: Optimal Values for Sigma on various facial features using a sigma step-size of 0.1.

	Training Data	Testing Data
R. Eye	2.6 and 2.8	2.5 and 3.1
L. Eye	2.4	2.8
Nose	1	1
Mouth	1	1.7

When comparing Figure 4.4 with Figure 4.1, we note the similar structure among the total summed squared image error, the variable radius complement average image error, and the total number of failures over the data set. This evidence supports our conclusion that when the summed squared image error is minimized, we achieve better results in that the strongest response in a filtered image r_1 is highly likely to be located on or close to the desired target.

4.3 Non-Spherical Gaussians

The shape of any given feature we wish to detect is likely to not be spherical (or circular given the projection of the subject onto a two-dimensional image). Until now, we have been working with and optimizing a spherical Gaussian in g_j. We know from Chapter 3.2 that the Fourier Transform of a Gaussian is a Gaussian. That property holds even if the Gaussian is not spherical. So, given a feature like the mouth which is longer (more horizontal) than tall (more vertical), we consider trying various shapes of Gaussians.

In Figures 4.5 - 4.8 we explore using non-spherical Gaussians through the 4 target features of this data set. To generate these results, we redefined

$$g_j(x) = \exp^{\frac{-||x_1 - x_{1,j}||_2^2}{2\sigma_1^2} - \frac{||x_2 - x_{2,j}||_2^2}{2\sigma_2^2}}$$

where σ_1 modifies the horizontal component of the Gaussian and σ_2 modifies the vertical component of the Gaussian.

Operating under the same criteria as we used in Chapter 4.2, we create our MOSSE filter using 512 images. Once the MOSSE filter has been applied to the remaining 2852 images, if the maximum pixel intensity in r_1 is with a 10 pixel radius (using the l_∞-norm), then we consider that a success. In the given figures, we plot σ_1 and σ_2 against the percent of successes. The step size for both σ's is 0.2.

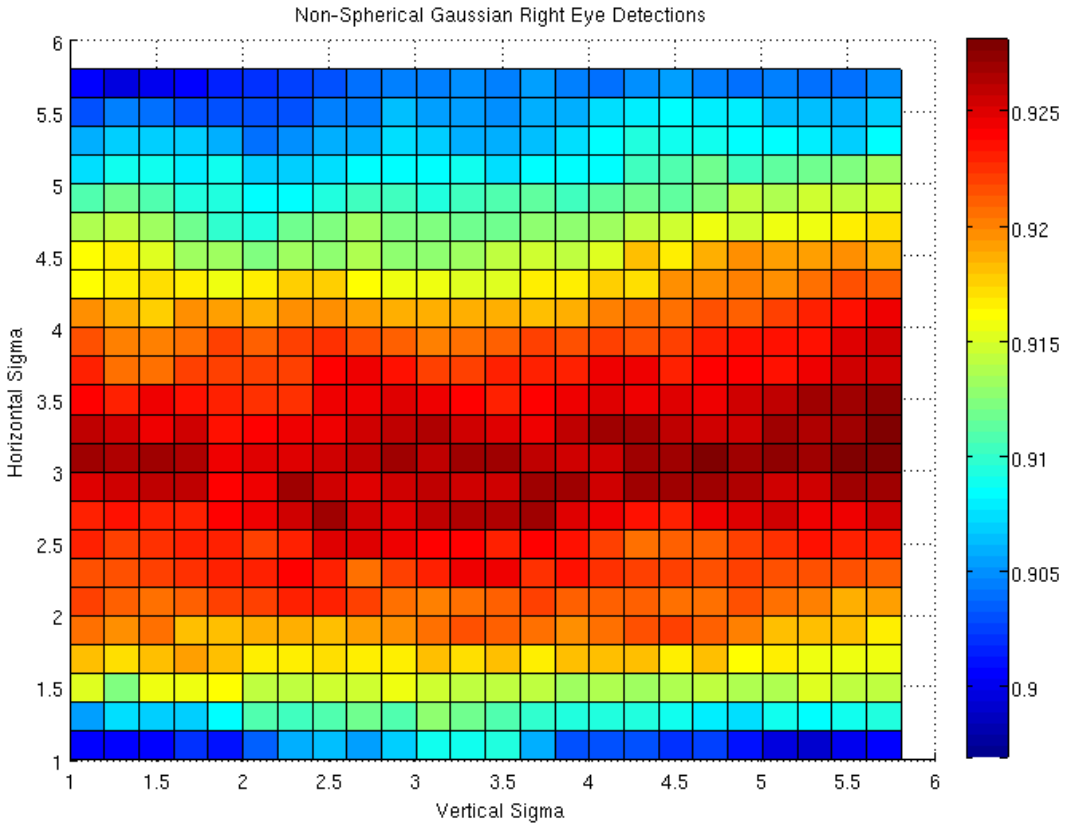

Figure 4.5: FERET non-spherical Gaussian right eye detection rate.

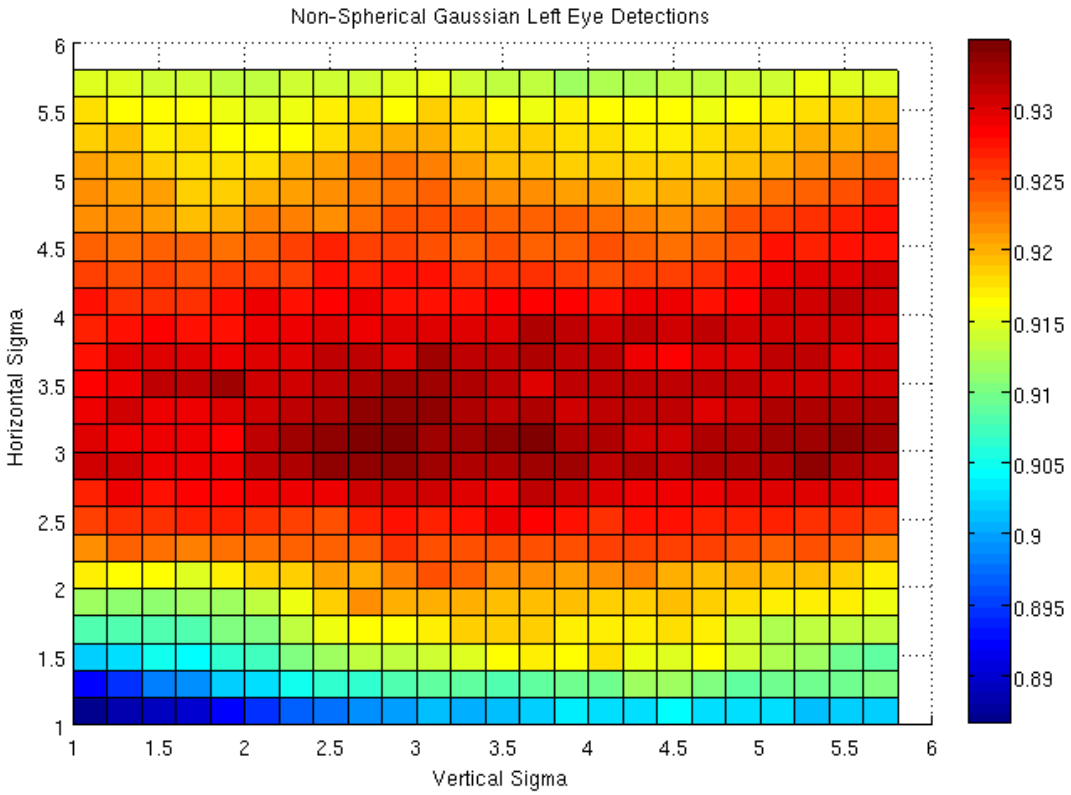

Figure 4.6: FERET non-spherical Gaussian left eye detection rate.

While the optimal values for the σ's are close to the values we found when using a spherical Gaussian, however they are not identical and as we will show in Chapter 4.9, they do yield more accurate results than the optimal spherical σ. In Table 4.2 we show optimal σ values for both spherical and non-spherical Gaussians.

Table 4.2: Optimal value of sigma for spherical and non-spherical Gaussians.

	σ_1	σ_2	Spherical σ
R. Eye	4.6	3	2.5 and 3.1
L. Eye	3.6	3	2.8
Nose	1	1	1
Mouth	1	2.4	1.7

For the rest of this chapter, unless otherwise stated, we will be performing future experiments using the optimal spherical Gaussian. Despite obtaining better results with the

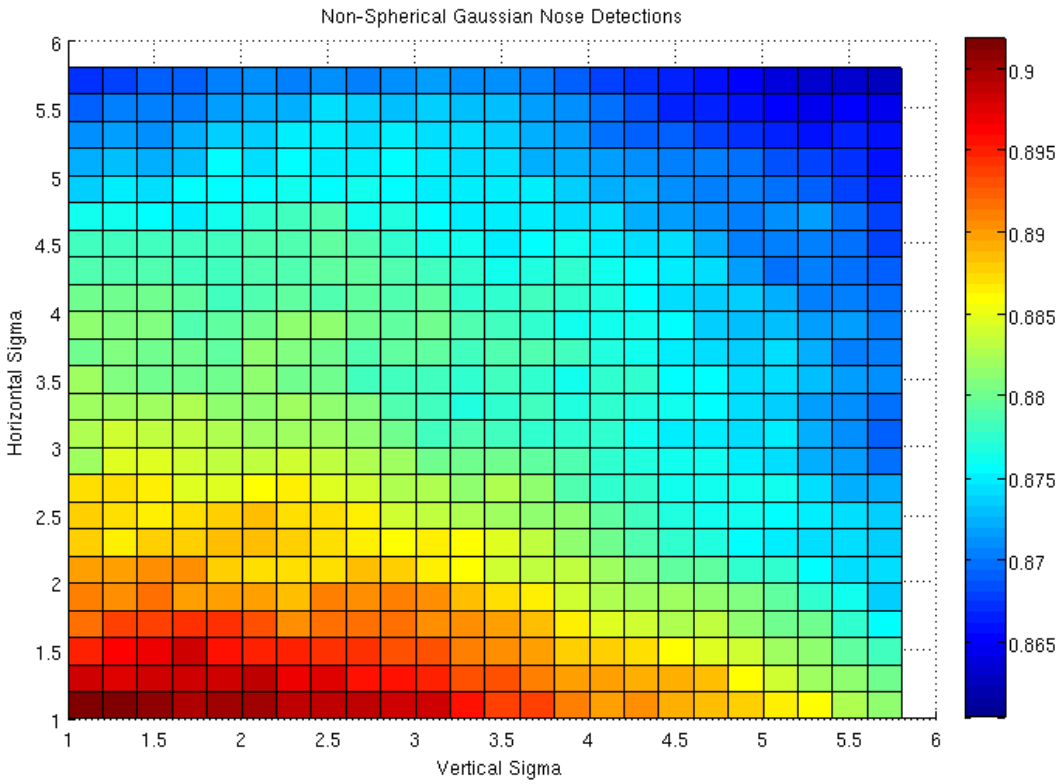

Figure 4.7: FERET non-spherical Gaussian nose detection rate.

non-spherical Gaussians (in all cases except the nose), we will sacrifice the small increase in performance for computational efficiency.

4.4 Analyzing the Response using PSLR

Once we compute our pixel response, we need to be able to evaluate criteria for determining detections. As we have seen, there are many local peaks in the response

$$r_1 = \mathcal{F}^{-1}\left(T \odot M^*\right).$$

Our goal is to determine which of these peaks (if any) are at a location that represents the pixel location of the desired target in the original test data t. We will start by examining the detection method used by Bolme [2].

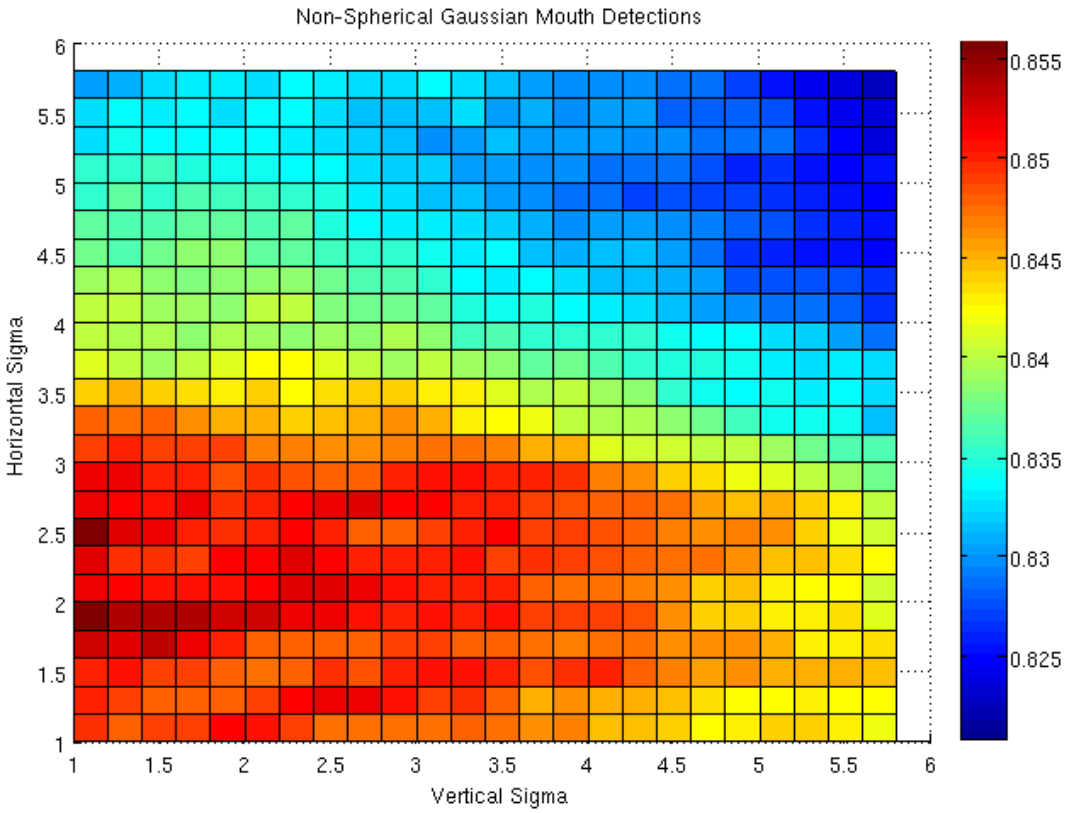

Figure 4.8: FERET non-spherical Gaussian mouth detection rate.

In pixel space we can examine the response signal and determine where in a given testing signal, t, is the feature we identified when creating our w's. One of the features of the MOSSE filters that we have discovered is that the total energy of the filter is dependent on σ and the frequencies in $W_i \odot F_i$. To equalize the effects in physical space, we normalize r_1 so $\min(r_1) = 0$ and $\min(r_1) = 1$ by the following:

$$\hat{r}_1 = \frac{r_1 - \min(r_1)}{\max(r_1)}.$$

Alternatively, we could also create g_j without the scalar term, so

$$g_j(x) = \exp^{\frac{-||x-x_j||_2^2}{2\sigma^2}}.$$

From this point on, we will assume one of the above methods has been used so $\max(r_1) = 1$ and $\min(r_1) = 0$. To analyze this signal we use a Peak-to-Side-Lobe Ratio (PSLR) test [3]. This method of detection is applied to $\hat{r_1}$ in physical space. To perform this test, we start by defining $Q(i,j)$ to be the pixel intensity at a pixel location (i,j) and the average

$$\overline{Q}(i,j) = \frac{1}{N-1} \sum_{\substack{(m,n)\in I_{i,j} \\ (m,n)\neq(i,j)}} Q(m,n)$$

where $I_{i,j}$ indexes pixels in $B_{i,j}$ and N is the number of pixels in $B_{i,j}$.

Then,

$$PSLR(i,j) = \frac{Q(i,j)}{\overline{Q}(i,j)}.$$

And the location of the maximum PSLR is given by

$$(i^*, j^*) = \underset{i,j}{\operatorname{argmax}}\, PSLR(i,j).$$

From the structure in r_1 we will only compute the PSLR values on pixel location which are local peaks. Furthermore, we will declare (i^*, j^*) to be a true detection if the distance between the pixel optimal pixel location and the ground truth for a given image is within a certain distance of each other.

4.5 Better PSLR Calculations

One issue with the original method for computing the PSLR (as described in the previous section) is that it favors small values of σ since Gaussian with small σ's have a large slope when measured from the peak. While it is ideal to create our w_i filters using Gaussians with small σ, the ramifications of mislabeling a point of interest, even by one or two pixels, are such that all response signals would be changed. In order to help normalize the PSLR across various values of σ, and thereby making it easier to select points of interest, we modify the algorithm to compute the peak by taking an average centered on the local maximum instead

of taking the peak value. To state the computation for the Modified PSLR (PSLRm), we first define $I_{i,j}$ to be the index set of $B_{i,j}$ and $K_{i,j}$ to be the index set of $B'_{i,j}$ where $B'_{i,j} < B_{i,j}$. Then

$$\overline{N}(i,j) = \frac{1}{K} \sum_{(m,n)\in K_{i,j}} Q(m,n)$$

and

$$\overline{D}(i,j) = \frac{1}{I-K} \sum_{(m,n)\in I_{i,j}\setminus K_{i,j}} Q(m,n)$$

where K is the number of pixels in the small box and I is the number of pixels in the large box.

Now we can define the PSLRm as

$$PSLR(i,j) = \frac{\overline{N}(i,j)}{\overline{D}(i,j)}$$

where the location of the maximum PSLRm is given by

$$(i^*, j^*) = \underset{i,j}{\operatorname{argmax}}\, PSLRm(i,j)$$

This choice of $B_{i,j}$ gives a radius around the maximum that would contain a significant portion of the Gaussian, assuming the response is ideal. If their response is not ideal, this width will ensure a low ratio.

Once again, instead of running the PSLR (or modified PSLR) computations over every pixel in r_1, we will only select local peaks as the center point for PSLR computations. Our algorithm for determining local peaks is described in Chapter 4.7. After running the PSLR algorithm over all local peaks, the pixel location which produces the maximum PSLR value will be declared the target location.

We have combined the results of the PSLR and modified PSLR computations for all 4 facial features (right eye, left eye, nose, and mouth) in Figure 4.9 using $c = 10$ while varying s from 0 (which gives us the normal PSLR computation) to 4.

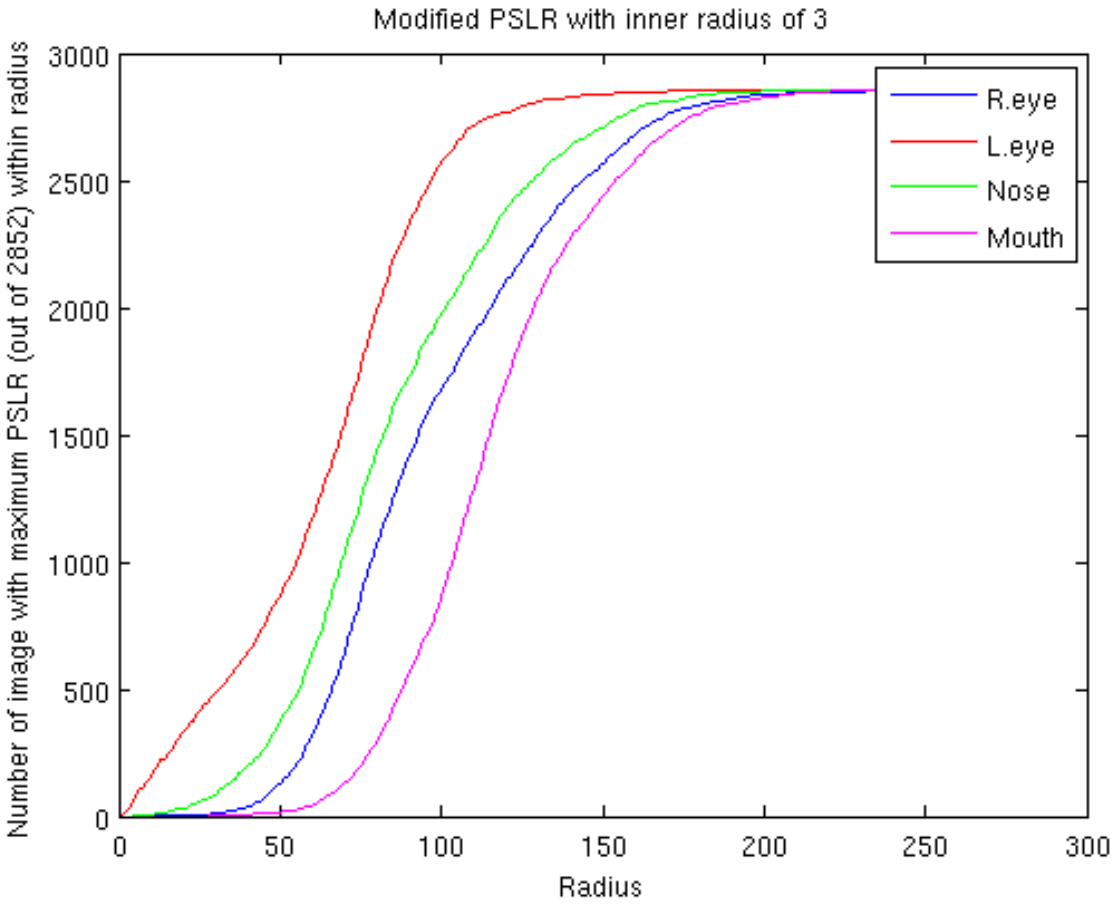

Figure 4.9: FERET modified PSLR results.

4.6 Maximum Value

Unfortunately, both the PSLR and modified PSLR detections when examined over the entire pixel response r_1, produce mixed results and can be computationally expensive. In [3], we note that the PSLR computation was used after a tracker had already identified the region in an image that contained the target feature. The PSLR detector at that point was able to identify the exact (or close to exact) location of the desired target within that window fairly easily, as the windowed response distribution favored this method of computation. However, examining the entire pixel response (not-windowed), we observe much stronger ratio responses on the false positives than true positives.

The PSLR does have some advantages over, e.g. a max intensity detection criterion, at least in smaller windows. For example, since it is a ratio, it is not dependent on the pixel intensity. When comparing the pixel response generated by applying the MOSSE filter to t_1 and t_2, we note the maximum pixel intensity is not necessarily the same. By using the PSLR we can easily select a threshold value and declare the desired target is in the test image if that PSLR value is above a certain number. Likewise, if the PSLR value is not above a certain number, we declare the target is not in the test image.

However, in this thesis, we explore the more difficult problem of examining the pixel response over the entire image, not a windowed area where the desired target is known to exist. To make this process slightly easier, we take advantage of the fact that every image in our testing and training data sets contain the desired features (meaning every image will contain a right eye, left eye, nose, and mouth). So for detection purposes, we will declare the location of the maximum pixel intensity to be the found target. When we test r_1 for successes (or failures) in matching the filter's found target location with the ground truth data, we will give the results as a function of distances from the ground truth using the l_∞-norm. This will allow us to see the effects of our algorithms without being too concerned with errors in ground truth data.

4.7 Gaussian Iterative Method (GIR)

As we have seen in Figure 4.3 the pixel response, r_1, has numerous false peaks. In order to get the highest correct detection rate possible, we need to improve the strength of the true peak and decrease the strength of any false peaks. To do this, we look at several processing techniques on the pixel response data.

In our new algorithm for detection, we try to take advantage of knowing the ideal desired size and shape of the response for positive target points will be a Gaussian of the same size and shape we used in creating g_j for the MOSSE filter.

Since our choice of a mask for the original MOSSE filter was comprised with Gaussians of size σ, we are training the filter to produce Gaussians of size σ as outputs. By examining the first-order response signal, R_1 or r_1, we notice the signal is made up of a sum of Gaussians of various scales (both magnitude and size) and noise which does not have a Gaussian distribution. We know by the manner of which the MOSSE filter was created that only areas of the signal in r_1 with a Gaussian distribution can be detected actions.

Our eventual goal will be to identify the part of the response signal r_1 which is comprised of a sum of Gaussians. To start doing this, we identify local peaks in r_1 by creating a gradient map.

1. Start with the $(i, j)th$ component of r_1.
2. Identify all pixels, $k_1, k_2, ..., k_n$ with a radius of 1 from the $(i, j)th$ component in the l_∞-norm.
3. If $(i, j) - k_m \geq 0$, $\forall m \in [1...n]$, then the $(i, j)th$ pixel is a local maximum.

Now define the set s' associated with a r_1 to be the set of local maximums for r_1. Our mask filter w' is the desired response in physical space and is given by

$$w'(x) = \sum_{j|x_j \in s'} g_j(x).$$

Now we can start to identify the sum of Gaussians in r_1 by doing a correlation between R_1 and W', where

$$W' = \mathcal{F}(w')$$

Start by mapping r_1 to frequency space to obtain R_1. Then we compute a second response signal

$$R_2 = R_1 \odot W'.$$

Taking the inverse fast Fourier transform of R_2 will give us the response signal r_2, in pixel space. Again, we normalize this response signal so

$$\min(r_2) = 0, \ \max(r_2) = 1.$$

To compute the correlation between the actual response r_1 and the expected response r_2 we compute

$$v = r_1 \odot r_2$$

where $\odot$ is a pointwise multiplication. To normalize the results across data sets we normalize v by setting $\hat{v}$ where $\min(\hat{v}) = 0$ and $\max(\hat{v}) = 1$.

By using the maximum pixel response location as described in Chapter 4.6, we then compute the number of failures as a function of the radius away from the given ground truth data and display the results in 4.10.

Unfortunately, we did not see any improvement in the detection rate by using GIR using the FERET data set. A closer examination of the r_1 (without the GIR method applied) around the ground truth location reveals that in a non-trivial data set, such as FERET, the actual output is not a Gaussian of the same size and shape as the Gaussians used in the creation of g_j. Instead, there are often summed Gaussians, where the filter produced a strong response on the center of the target feature, and a strong response within a few pixels of that (such as left of the center of the eye and to the right of the center of the eye).

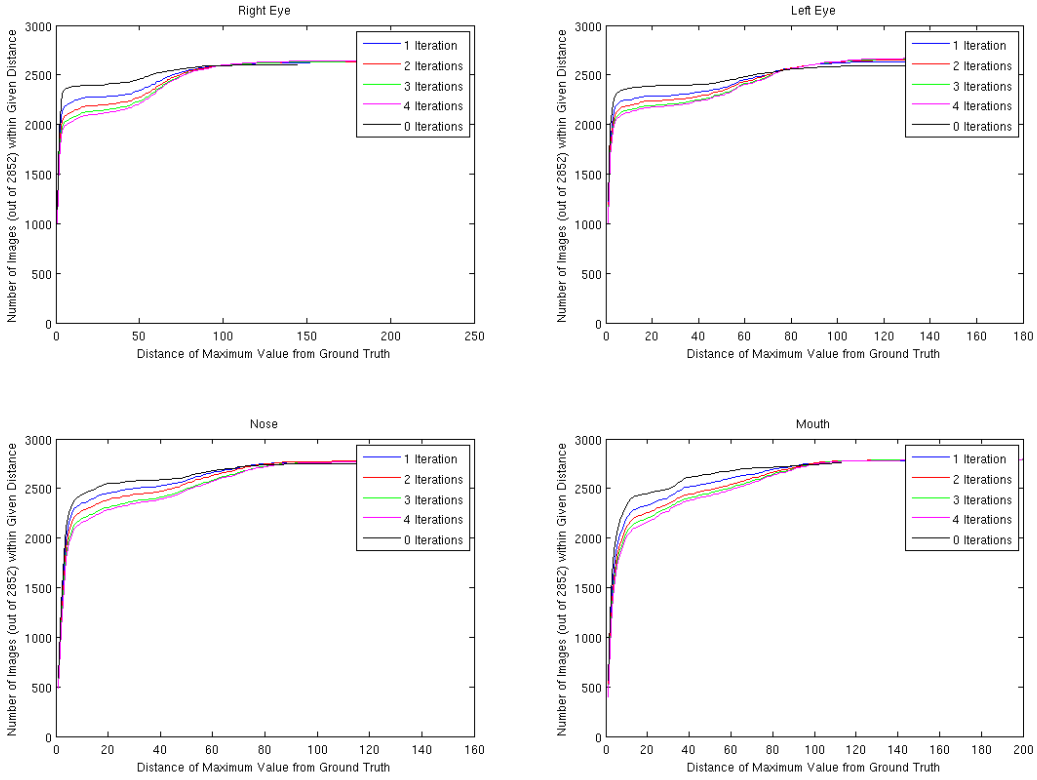

Figure 4.10: FERET GIR successes.

As these Gaussians are summed, the result is not Gaussian. Other common output features include a response that appears Gaussian, but elongated. For example, the mouth often has a horizontally elongated or rotated Gaussian (as that is the common appearance of the mouth - it is more long than tall).

As we kept convolving the pixel responses with spherical Gaussians, we often lost the positive detections due to their non-uniform structure. At the same time, there are many false positives where the filter only responded at a single pixel location which produced spherical Gaussians that we were training on. These false positives remained near the same pixel intensity while the true positives decreased in intensity.

4.8 Multi-Sigma Products

As we determined in Chapter 4.1 and 4.2, the value of σ used in the creation of w_i is critical to reduce the number of false detections. However, if we were to examine the particular images that failed to have the maximum response located near the ground truth data, would we find that the same images fail as we vary σ? And if so, would they have false detections in the same location with the same pixel intensity?

An examination of the MOSSE filters in Figure 4.11 using a series of different σ values will show us the answer to the latter question is 'no'. The size and shape of facial features in the MOSSE filter has changed as well as relative pixel intensities.

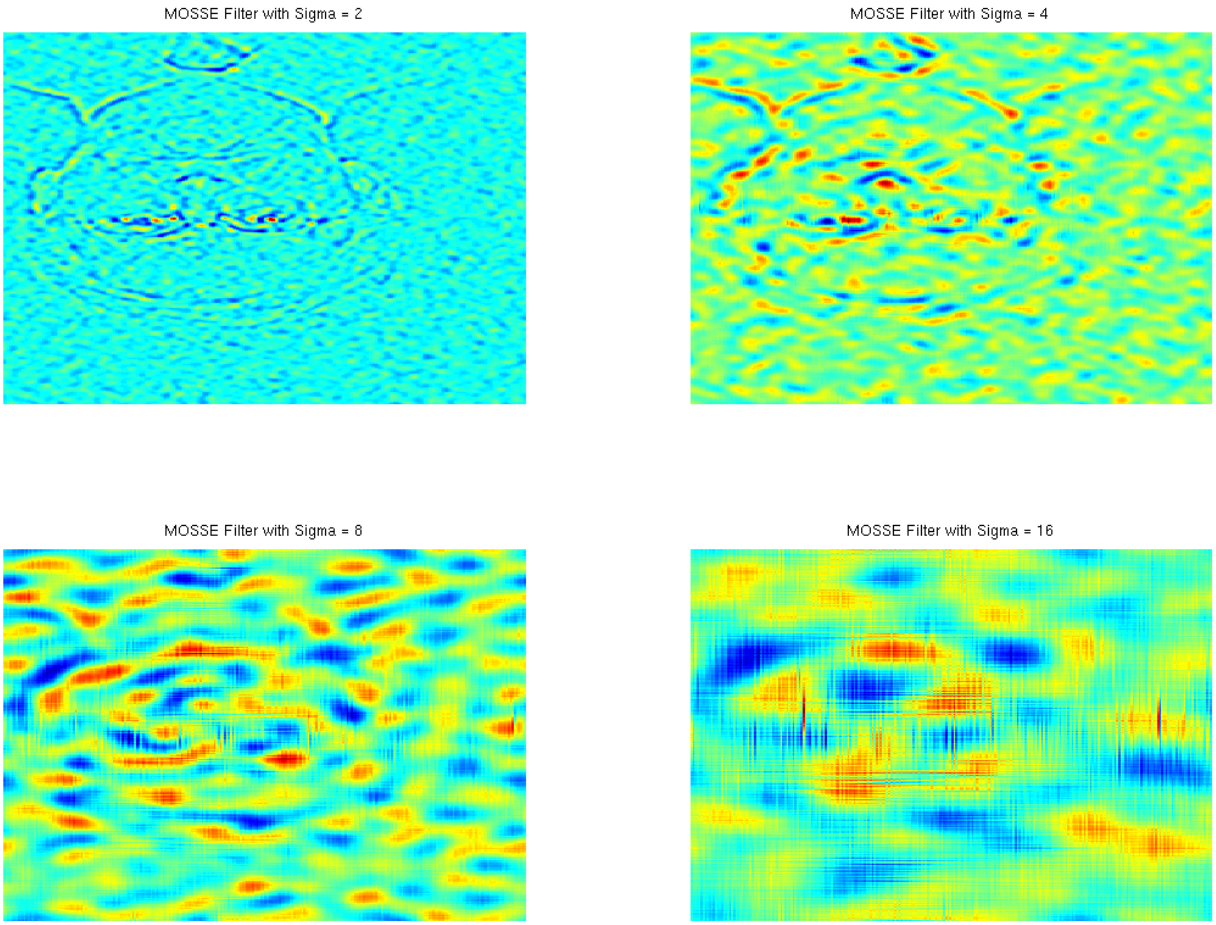

Figure 4.11: FERET MOSSE right eye filters.

In an ideal world, no matter what the value of σ is, we expect the greatest response to be located on the target feature and every false detection would have a pixel intensity less than that of the target. Obviously, that is not the case.

However, we can use the results from MOSSE filters with different values of σ to our advantage. Consider this scenario: Given σ_1 and σ_2, such that $\sigma_1 \neq \sigma_2$, we generate a MOSSE filter with each σ value, then apply the MOSSE filters to a single test image t to get two pixel response, r_1 and r_2 respectively. We examine r_1 and find it has a positive true detection at $(40, 96)$ with pixel intensity of 1 and a fairly strong false detection at $(80, 140)$ with pixel intensity of 0.8. By examining r_2 at the same coordinates, we find the positive true detection only has a pixel detection of 0.9 while the false detection has a pixel intensity of 1. If we were to test r_2, we would classify that as a failure. However, if we compute the Hadamard Product of r_1 and r_2 by performing a pixel-wise multiplication between the two, the result would have a pixel intensity at the true positive as 0.9 while the false positive would be lowered to 0.8. While the true positive value might slightly lower through these products, we expect to see a dramatic decrease in the intensity values of the false positives to the point where we would obtain a true positive detection.

We have performed 2 tests with this method. In both cases, we used a training data set of 512 images and a testing data set of size 2852. For the first test, we generated and applied MOSSE filters created using σ values of $2, 4, 8$, and 16 and took the Hadamard Product of the pixel responses generated by using $\sigma = 2$, $\sigma = 2$ & 4, $\sigma = 2, 4$, & 8, and $\sigma = 2, 4, 8$, & 16. These results are displayed in Figure 4.12.

For the second test, instead of using $\sigma = 2$, we replaced this value with the optimal sigma's per facial feature that we found in Chapter 4.2. The results from this test can be found in Figure 4.13. Note that the blue lines represent just using the optimal value for σ (meaning we do not take a Hadamard Product with other σ values). By comparing these two sets of plots, we can conclude that if we do not know what the optimal σ is, this method can improve results by a few percent.

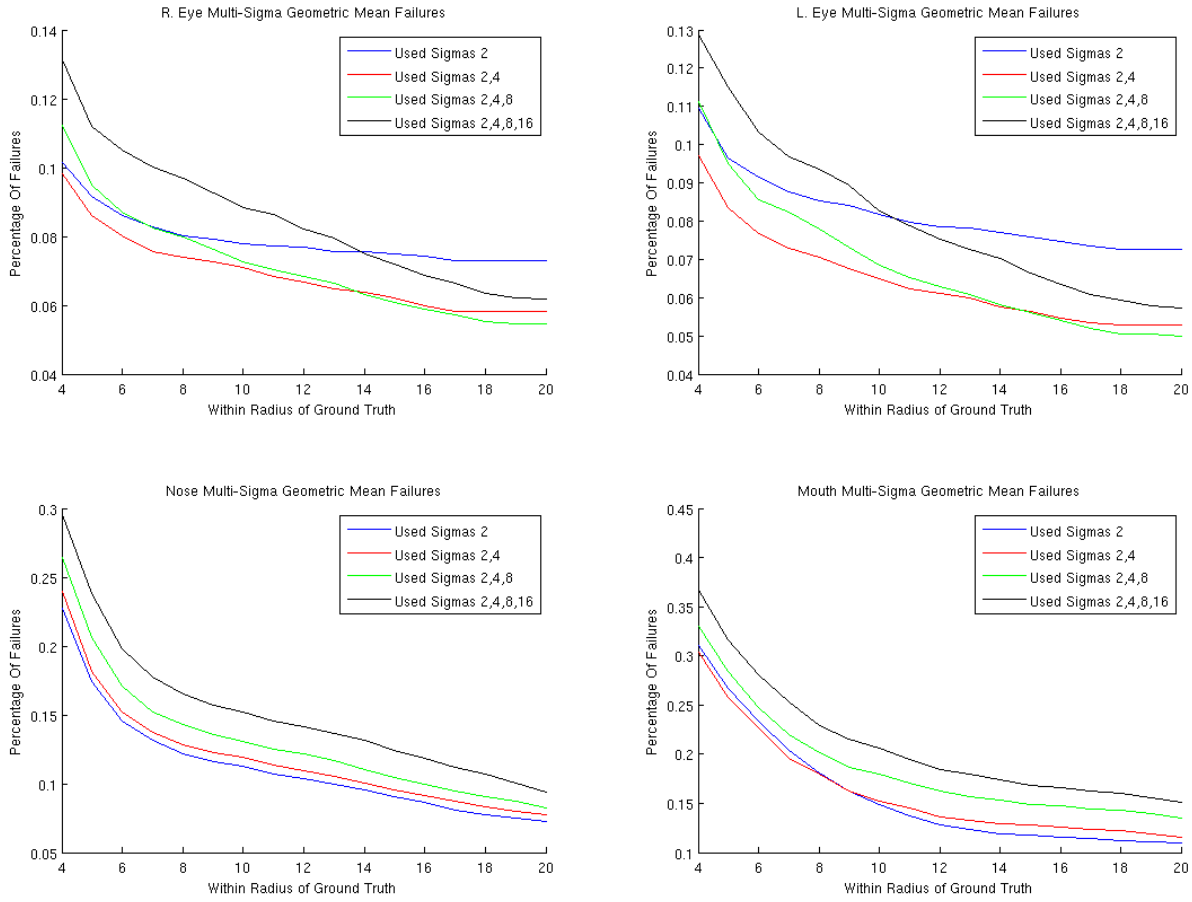

Figure 4.12: FERET multi-sigma product detection results.

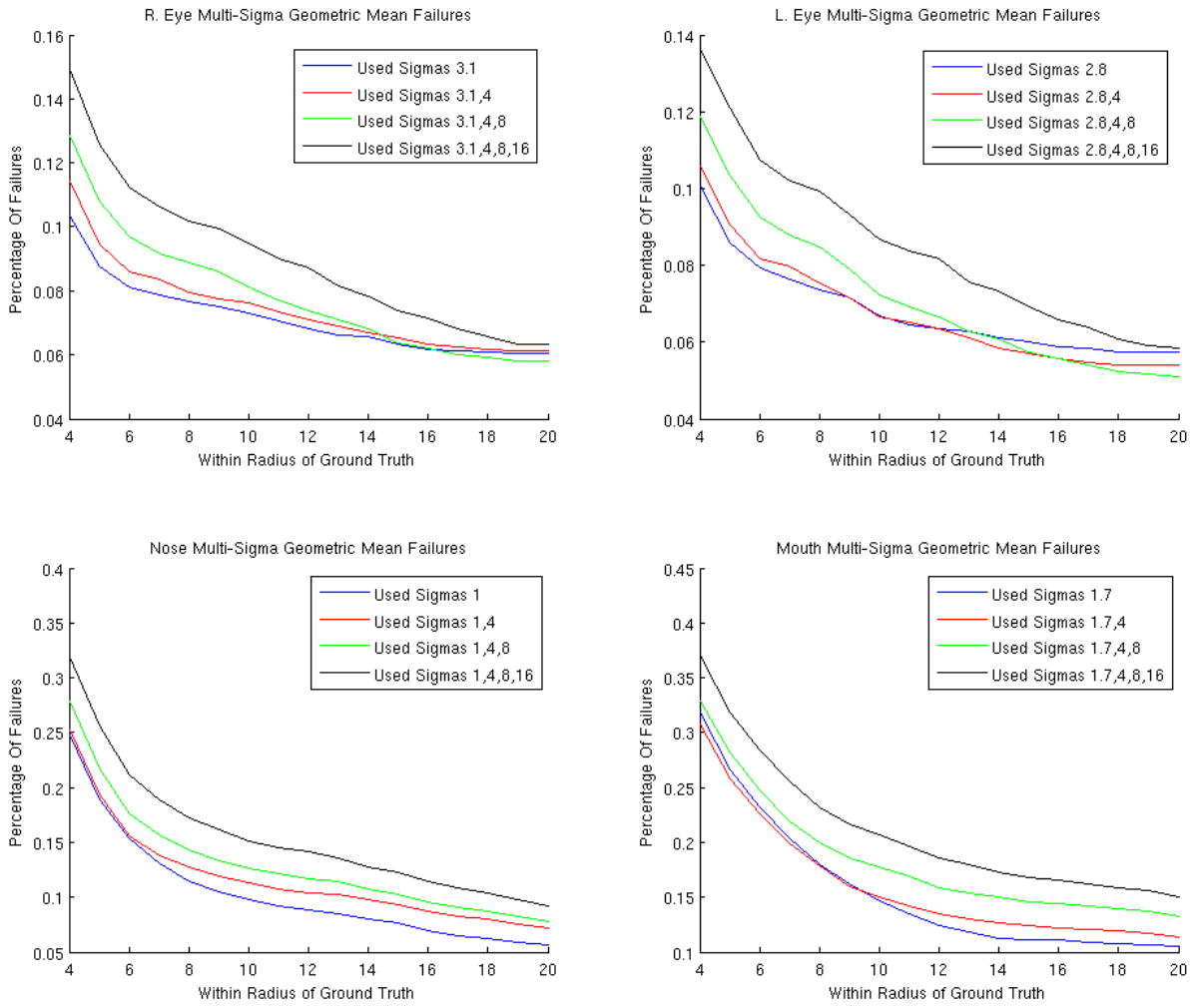

Figure 4.13: FERET multi-sigma product with optimal sigma detection results.

When comparing these results with Figure 4.1, we can start to understand this behavior. As we included responses generated with $\sigma = 8$ and 16 in the Hadamard Product, the variable radius complement error shows us there are strong false detections. As we use an optimal σ value, we are increasing the intensity of (at least a few) false detections. However, as we only include $\sigma = 4$ with our optimal σ (or $\sigma = 2$) the amount of error that is not around the target feature is still relatively small. In the case where $\sigma = 2$ our initial hypothesis remains correct, we are essentially averaging out the false detections while keeping the true positives. However when we use the optimal σ the location of false detections within an image is not unique against σ. So in effect, we are slightly increasing the intensity of false detections.

4.9 FERET Conclusions

In this chapter, we have been running simulations that involve creating a MOSSE filter using 512 images and testing the filter on the remaining 2852. This number was selected based upon the prior work done by Bolme [1]. It was concluded that around 512 samples, the MOSSE filter obtained optimal results. For completeness, and to show how our new techniques works with fewer training samples, we ran various simulations with MOSSE filters created from 4, 8, 16, 32, 64, 128, 256, and 512 images. The results are displayed in Figures 4.14 - 4.17.

In all 4 features, we see marginal improvements over the non-optimal spherical σ in which $\sigma = 2$ by using some of our new methods such as Optimal σ, Non-Spherical Gaussian, and Multi-Sigma Geometric Means. Results from Gaussian Iterative Response, the PSLR, and modified PSLR were not included as the accuracy of detections did not improve on any feature with any number of training sizes. We also see that once we reach 'enough' images in the training size, the use of a non-spherical optimal Gaussian outperforms all the rest.

In summary, we have gained several insights about MOSSE filters, how to construct them, and the outputs obtained by applying the filter to a set of test images:

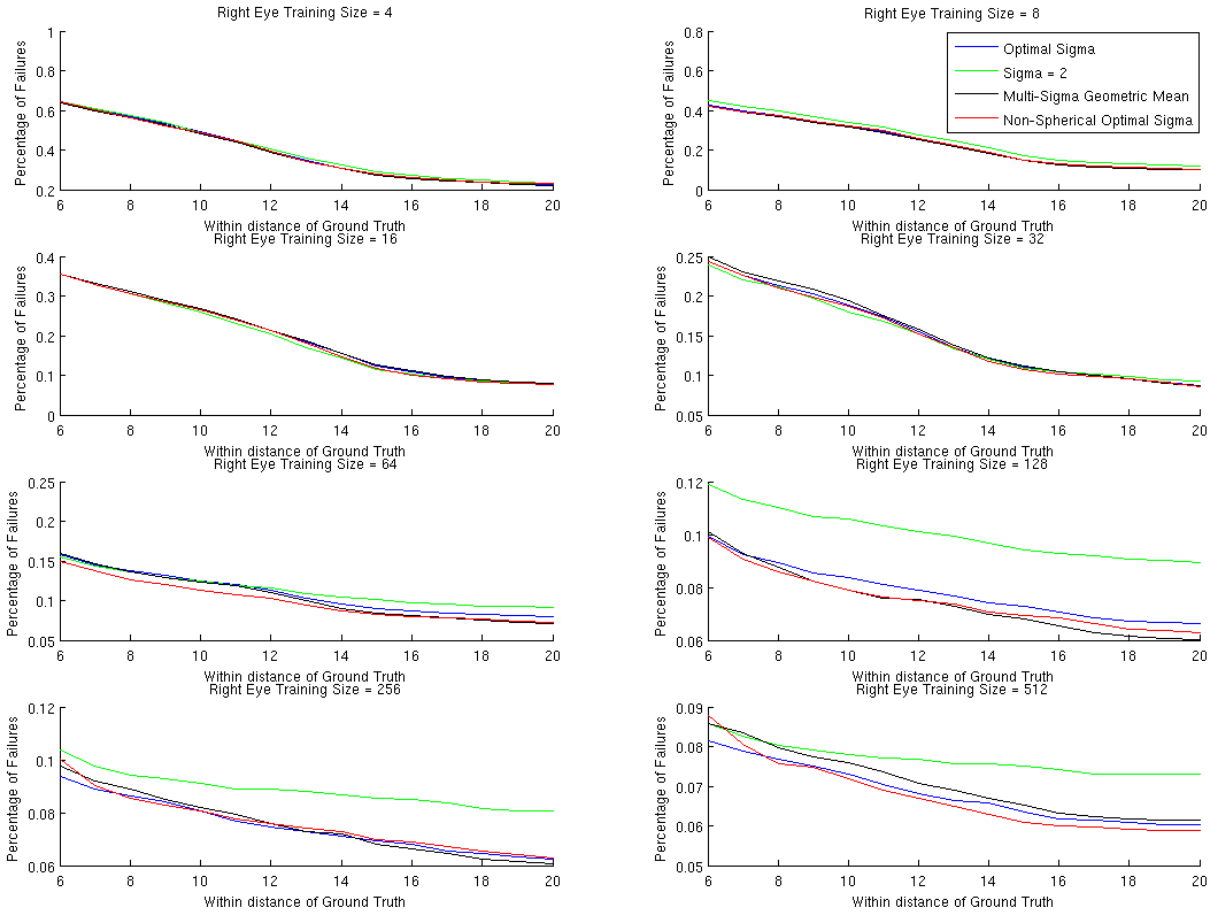

Figure 4.14: FERET right eye comparison of methods with various training sizes.

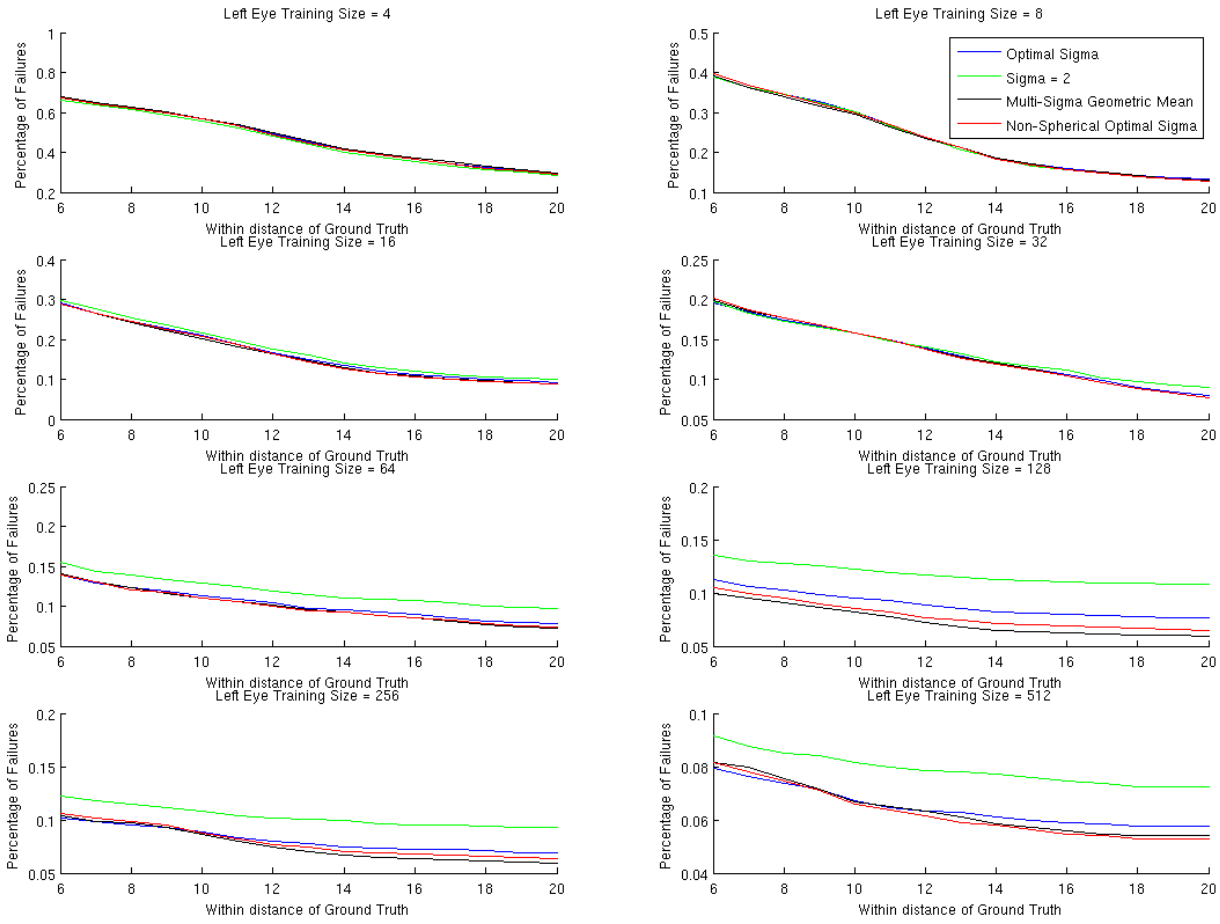

Figure 4.15: FERET left eye comparison of methods with various training sizes.

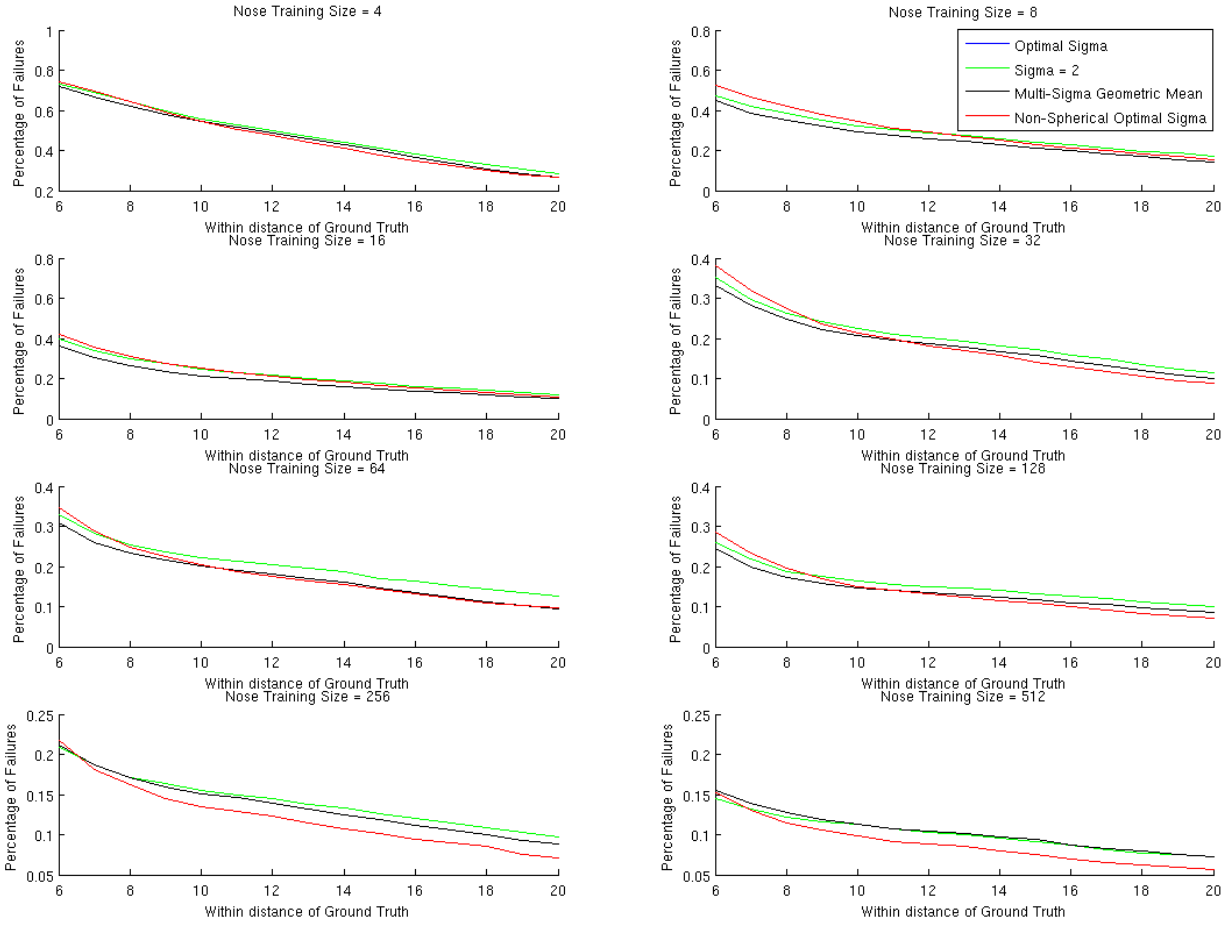

Figure 4.16: FERET nose comparison of methods with various training sizes.

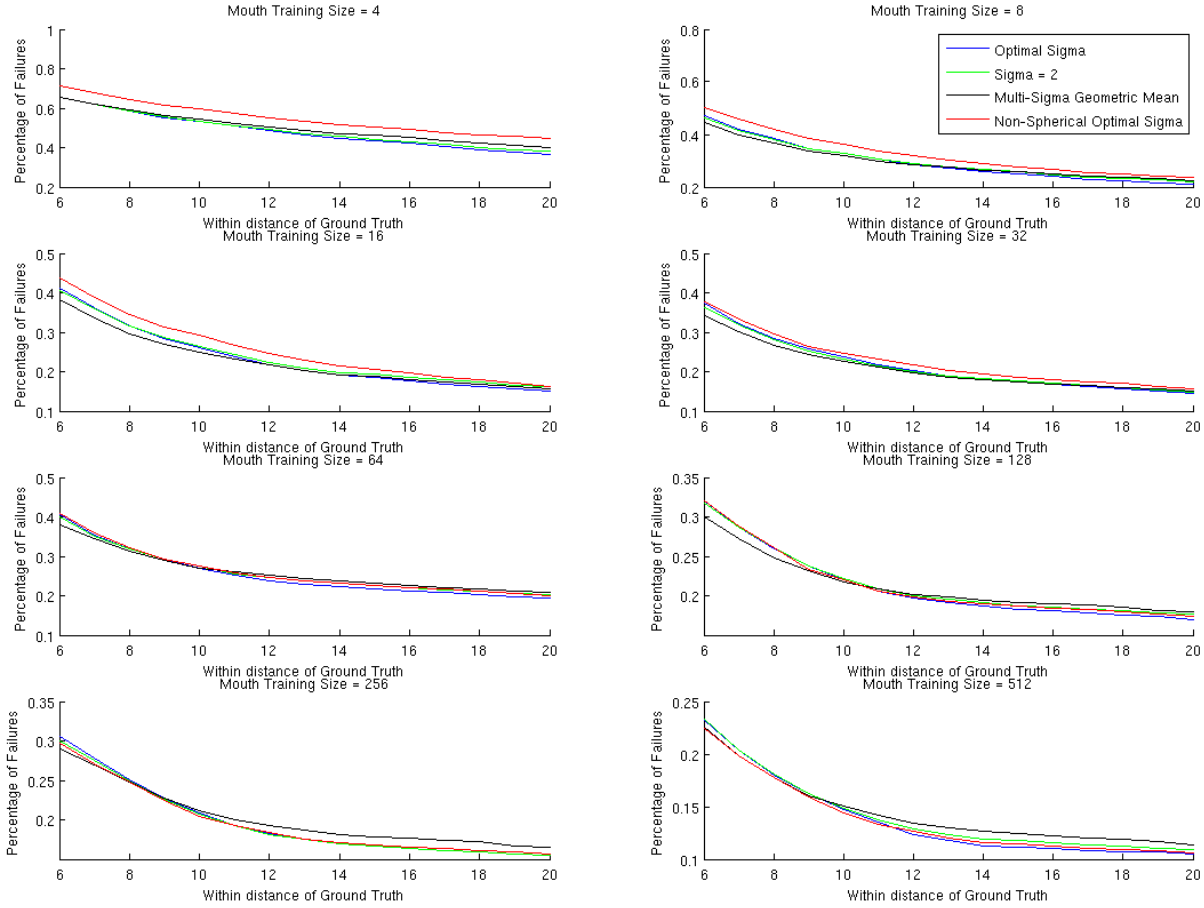

Figure 4.17: FERET mouth comparison of methods with various training sizes.

1. By minimizing the sum of squared error, we can find the value(s) of σ which optimize filter performance for a given target feature. In addition, by examining the error in various parts of the set of pixel responses, we have determined that the majority of the error is coming from false peaks, or locations where the filter response to features in the image which are not directly related to the desired feature.

2. By examining the PSLR and PSLRm, we have determined that although the PSLRm value is high on the closest local peak to the ground truth data, in most cases there is at least one false peak in the pixel response that has a higher PSLRm value. In 2-D (image) data where the entire data set is known to contain the target feature, it is easier and more accurate to use the location of the maximum pixel intensity as the computed feature location.

3. By performing the GIR algorithm and by subsequently examining the location around the ground truth for a given feature, we have concluded that the size and shape of the true peak is not the same size and shape as the Gaussian used to create w_i. Furthermore, smoothing responses, r_i, with a set of Gaussians will decrease detection accuracy.

4. By examining the pixel responses generated by using the Multi-Sigma Geometric Mean approach, we have concluded that the location and intensity of the false peaks varies as σ varies. If we do not know the optimal value(s) of σ to use for a particular feature, but rather, we know the approximate value(s) of σ we can use a Multi-Sigma Geometric Mean to reduce the intensity of false peaks while maintaining a strong true peak intensity.

We acknowledge that some of our techniques did not necessarily improve results on the data sets analyzed, and other techniques showed only marginal improvements. However, we conjecture that these algorithms may still prove to be valuable techniques. Further investigation is required to evaluate the utility of these new algorithms, e.g., on data sets

with much lower best classification rates. At this time however, we switch topics and start to explore the action detection problem.

Chapter 5

FIRST STEPS IN ACTION DETECTION

For applications to action detection in video, we propose to implement a 3-dimensional version of the MOSSE filter. The calculations extend in a natural way with the 2-D FFT being replaced by the 3-D FFT. In addition, the location of the action is now described by a 3-D Gaussian window, which again has a 3-D Gaussian FFT. In this setting we can explore videos as a whole instead of just operating on them frame by frame. By considering the spatial and temporal components together, we explore the use of MOSSE filter to create action detectors. In addition we might detect a sequence of similar spatially similar objects centered around a single point in space-time. In this chapter we explore the first steps of how to create an action detector and illustrate the challenges of implementing MOSSE in 3-D. Using the KTH data set, as described in Chapter 2. We note that despite this popular data set being available since 2003 [7], at the time of this writing, there are not any known published results for action detection on this data set.

The construction of 3-dimensional MOSSE filters follows from the steps outlined in Chapter 3.2, save for the preprocessing of the data. In the 2-dimensional FERET data, we set the mean of each image to 0, the standard deviation to 1, then applied a cosine window. This helped us to correct for lighting intensity differences and it filtered out high frequency artifacts. Given that our target points were located within the center of each FERET image, the cosine filter did not have much effect on the pixel values of the regions we were looking at. However, with the KTH data, we have numerous examples where an action occurs within the first or last few frames, or on the edges of one (or more) frames. The application of the cosine filter in these cases would make the structure of the action unrecognizable, even to the human eye. So as we explore the KTH data set, we will operate with the following condition: all videos used in the creation of the MOSSE filter have been preprocessed by

mean subtracting, setting the standard deviation to 1, and applying a cosine window, while all videos used in the testing data set have only been mean subtracted and set to a standard deviation of 1. Ideally, this will allow us to suppress high frequency responses by the MOSSE filter while still allowing us to have the ability to detect actions which occur on the sides of the video cube.

5.1 Selecting Sigma

We will start out as before in trying to determine the optimal value of σ to use in each MOSSE filter. We assume σ will depend significantly on the nature of the action. However, we modify the definition of g_j

$$g_j(x,y,t) = \exp \frac{-\left(\left(x-x_j\right)^2\right)^{\frac{1}{2}} - \left(\left(y-y_j\right)^2\right)^{\frac{1}{2}}}{2\sigma_1^2} - \frac{\left(\left(t-t_j\right)^2\right)^{\frac{1}{2}}}{2\sigma_2^2}$$

where x and y are the horizontal and vertical components which correspond to σ_1 and t is the temporal component corresponding to σ_2. By using this definition we are then able to create a non-spherical Gaussian that scales in time and space separately. When we take a temporal cross-section, we still have a circular Gaussian in the spatial domain, but a temporal cross-section will reveal an elongated Gaussian. This formulation allows us to exploit the ground truth of each action (as we described in detail in Appendix A).

We observe in these videos that the time an action takes (in frames) is much greater than the spatial extent of the action. By allowing our Gaussian to elongate in the temporal dimension, we can describe an action to the MOSSE filter over several frames while only looking at a small spatial region.

Also, by removing the scaling term in front of the exponential of $g_j(x)$ we remove the need to scale the pixel responses between 0 and 1. Since the intensity of the Gaussian is no longer a function of the value of σ used, all our trials involving varying σ will have the same relative scale and thus, can be directly compared.

With this in mind, we set out to find the optimal σ's, or the values (spatial and temporal) of σ, that will result in the fewest number of failures in the testing data set. For this experiment, we will only look at a subset of the videos. By the description of the data set in Chapter 2.2, we know that Scene 1 contains 25 people per action in an outdoor setting where the camera does not move. Scene 3 also contains 25 people per action in an outdoor setting where the camera does not move. The characterized difference between these scenes is that the actors/actresses are wearing different clothes. Since the other 2 scenes contain much more variability (changing camera and indoor setting), we only look at Scenes 1 and 3. For Figure 5.1 we trained 6 MOSSE filters, one for each of the 6 actions. Each MOSSE filter used in the 25 videos per action from Scene 3. The testing data is comprised from the 25 videos of the corresponding action in Scene 1. For the second run, we created the MOSSE filter using videos from Scene 1, then tested them on the same set of videos. The results are found in Figure 5.2.

In order to interpret these figures we first must consider the criteria to characterize detection. In order to call a detection a failure we adopted the following criteria. For the actions where the person is standing still (boxing, handclapping, and handwaving) we selected the top 3 local peaks (the method for detecting local peaks is given in Chapter 4.7). If the spatial coordinates are within 10 pixels (using the l_∞-norm) and the temporal coordinates are within a 3 frame distance (also using the l_∞-norm) of the ground truth for all 3 local peaks, then a true detection is said to occur, otherwise we call it a failure. For the actions where the person is not standing still (jogging, running, and walking) we selected the top 2 local peaks and performed the same test. For this experiment, these criteria were selected due to the number of examples of each action per video. The boxing, handclapping, and handwaving videos have many (5+) examples of the action per video while the jogging, running, and walking videos have, on average, far fewer.

In examining Figures 5.1 and 5.2 we see the actions in which the person is standing in one place (boxing, handwaving, and handclapping) are more insensitive to increases in

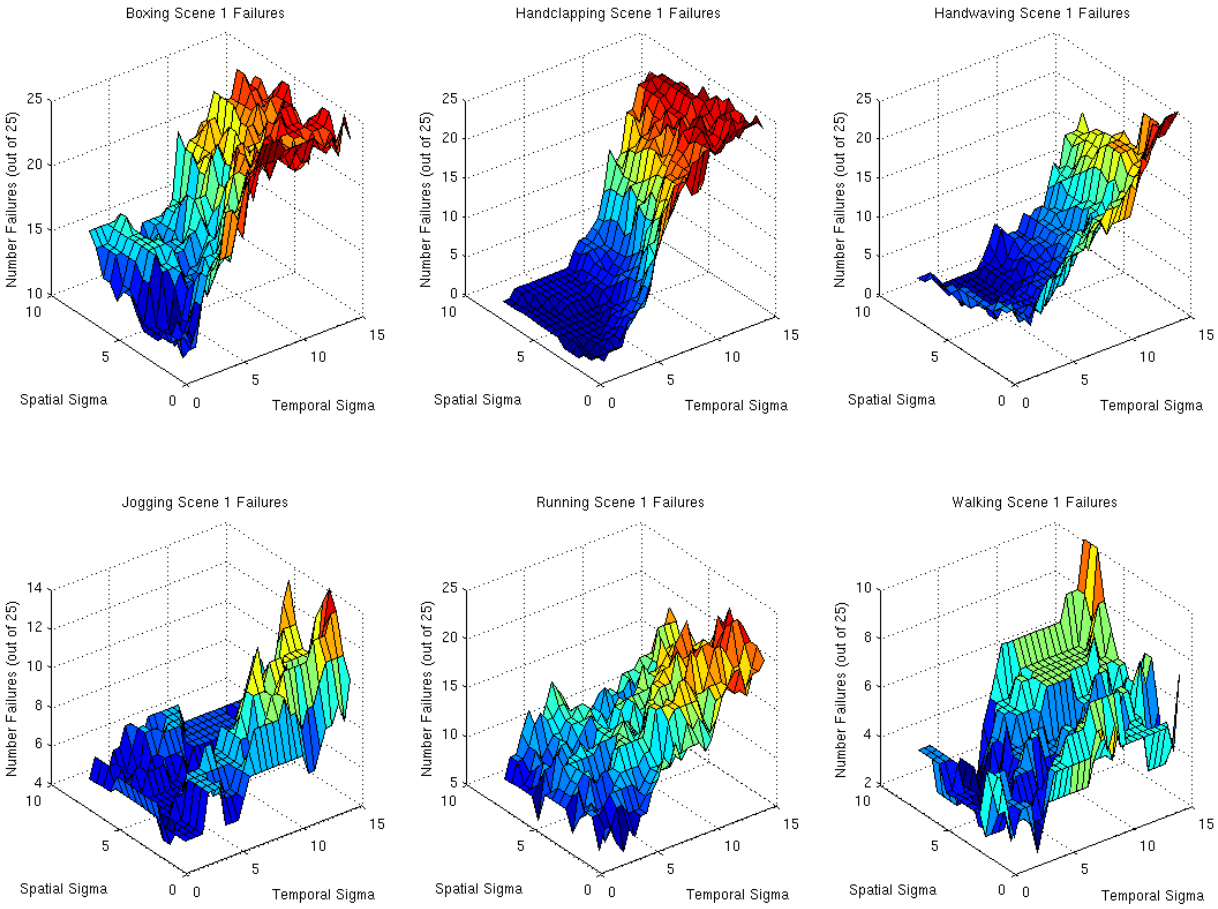

Figure 5.1: KTH Scene 3 created MOSSE Scene 1 test data failures.

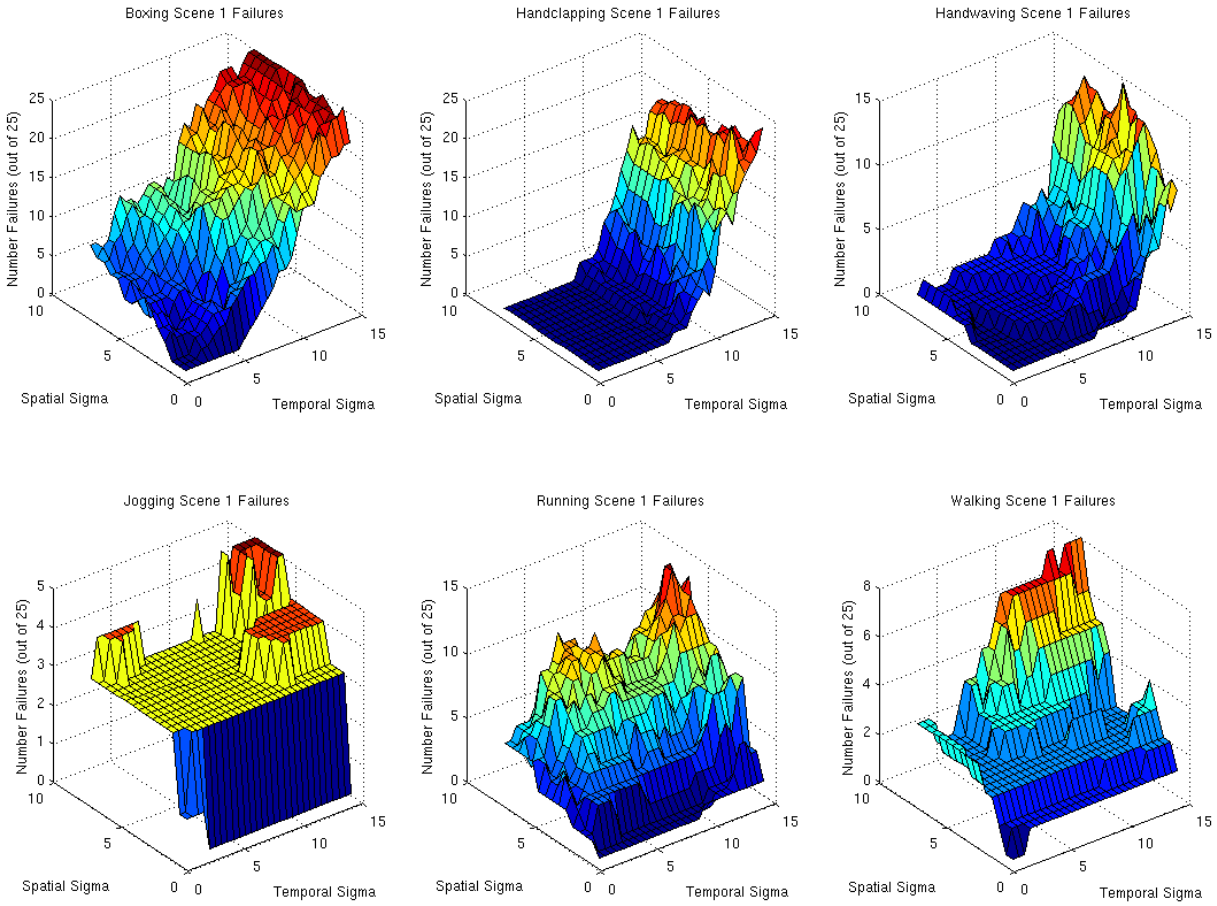

Figure 5.2: KTH Scene 1 created MOSSE Scene 1 test data failures.

the spatial domain compared to increases in the temporal domain. Similarly, actions in which the person is moving across the field of view (jogging, running, and walking) are more insensitive to increases in the temporal domain in comparison to the spatial domain.

As we see in all actions with both test data sets, the optimal results (where the number of failures is minimized) occurs when we set both the spatial and temporal σ to 1. Note that $\sigma = 1$ is the smallest value we can use to create a discrete Gaussian - any value smaller and the structure of the result would no longer approximate a Gaussian. Having this small a value of σ means that in the current configuration, the MOSSE filters work best as a pose detector rather than an action detector. We attribute this result to how ground truth data was selected (as described in Appendix A). Each ground truth point identifies the action based on a pose that is unique to the given action. As such, a $\sigma = 1$ value corresponds to the MOSSE filter heavily weighting frequencies contained in a very small windowed region around the ground truth point. In the case of a person running, the ground truth is identified on the tip of the back foot when the back leg is parallel to the ground. Since this particular pose does not occur when a person is jogging or walking, the MOSSE filter searches each test video for a black tipped foot that is pointed down. Virtually no temporal information is contained in the filter, nor is any other characteristic associated with running. With this in mind, we predict that a video with a person standing still with one leg raised such that the toe was pointed at the ground would also be identified as running.

This example illustrates challenges associated with action recognition using the KTH data set. Each video contains several examples of the given action. By looking at the ground truth data in Appendix A we see that the number of examples of an action occurring in a video varies, even within the action itself. With this in mind, once we have a pixel response, what criteria should we put into place that will allow us to go from the pixel response to the number and location of detected actions?

5.2 Negative Sample Training

Since the KTH data set is closed, i.e., there are a finite number of human actions (6 in particular) that we are trying to recognize, we can improve filter results by training on negative samples, or videos containing actions which are not the desired, target action. Let us consider 2 similar actions, say running versus walking. When we create our MOSSE filter on people running we are weighting frequencies contained in a windowed region around the pixel representation target point (in this case the tip of the back foot when the back leg is parallel with the ground) strongly, while at the same time weighting the frequencies that comprise the non-windowed pixel representation region of the data cube to zero. In this conceptualization, our filter is neutral to other actions. To elaborate, even though the MOSSE filter is weighted strongly on the frequencies that surround the foot with a parallel leg, it does not give any weight to the action of boxing. So when the walking MOSSE filter is tested on the boxing video

$$R_1 = \frac{\sum_{i=1}^{n} (F_i^* \odot W_i)}{\sum_{i=1}^{n} (F_i^* \odot F_i)} \odot T$$

any frequencies that are part of the boxing action which are not part of the small constant weighted extras from the running data cubes will partially respond to the filter. When the action of the MOSSE filter does not match the action of the test video, yet satisfies the detection criterion, the result is deemed a false positive.

To help with this problem, we will introduce negative samples into the creation of the MOSSE filter. A negative sample, in this context, refers to a video of an action other than the action we desire to detect, which is included in the creation of the MOSSE filter. To include a negative sample we set $w_i = 0$ for all video we use for negative samples.

Formally, if we define the set h to be the set of all signals h_k where $k \in \{1, 2, ..., z\}$ and z is the number of negative samples to include, then the MOSSE filter is defined as

$$M^* = \frac{\sum_{i=1}^{n} (F_i^* \odot W_i) + \sum_{k=1}^{z} (H_k^* \odot 0)}{\sum_{i=1}^{n} (F_i^* \odot F_i) + \sum_{k=1}^{z} (H_k^* \odot H_k)}$$

where $H_k = \mathcal{F}(h_k)$, H_k^* is the conjugate of H_k, and n is the number of desired action signals.

In Figures 5.3 - 5.5 we show 3 basic statistics (maximum pixel intensity per frame, average pixel intensity per frame, and standard deviation per frame) from when a MOSSE filter was created using walking videos and then the pixel response was computed by applying that MOSSE filter to various other actions. We used a leave one out method of testing in that the MOSSE filter was created by using the videos from 24 persons from the target action (a total of 120 positive samples). The test videos are from the person not used to build the MOSSE filter. Additionally, we included a number of negative samples in the MOSSE filter. In each figure, the statistics were plotted for when no negative samples were used, when 8 negative samples were used from each of the other 5 actions (for a total of 40 negative samples), and when 16, 32, and 64 negative samples were used from each of the other 5 actions (for a total of 80, 160, and 320 negative samples respectively).

We have investigated the maximum pixel intensity and standard deviation and after running many trials we cannot find a correlation between the average pixel intensity value and a positive (or false positive) detection. We observe the strongest pixel intensity value (and standard deviation) when we do not include any negative training samples, but do not view this as a problem. Our goal for detection is to be able to identify the true detections from the false positive detections. By making that determination by using a test such as PSLR, or some other test which involves measuring the declared true pixel response versus another value within the same test video, we conclude our goal should be to decrease the false detections at a faster rate than we decrease the true detections.

If we examine Figure 5.5 in which our MOSSE filter was constructed using the same action as the test video, we notice that by including 8 negative samples in the MOSSE filter, we still have a lot of structure at the locations corresponding to ground truth data. However, in Figure 5.3 and Figure 5.4, we notice that we lose most of the structure by including 8

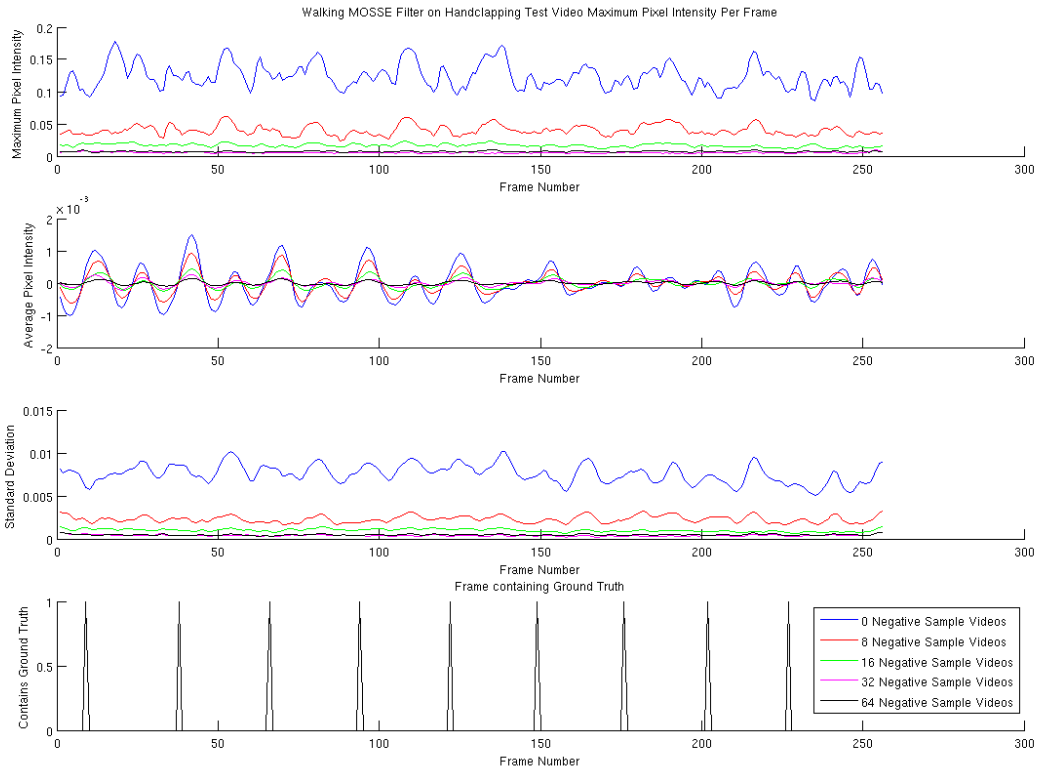

Figure 5.3: Statistics on walking MOSSE filter applied to handclapping subject.

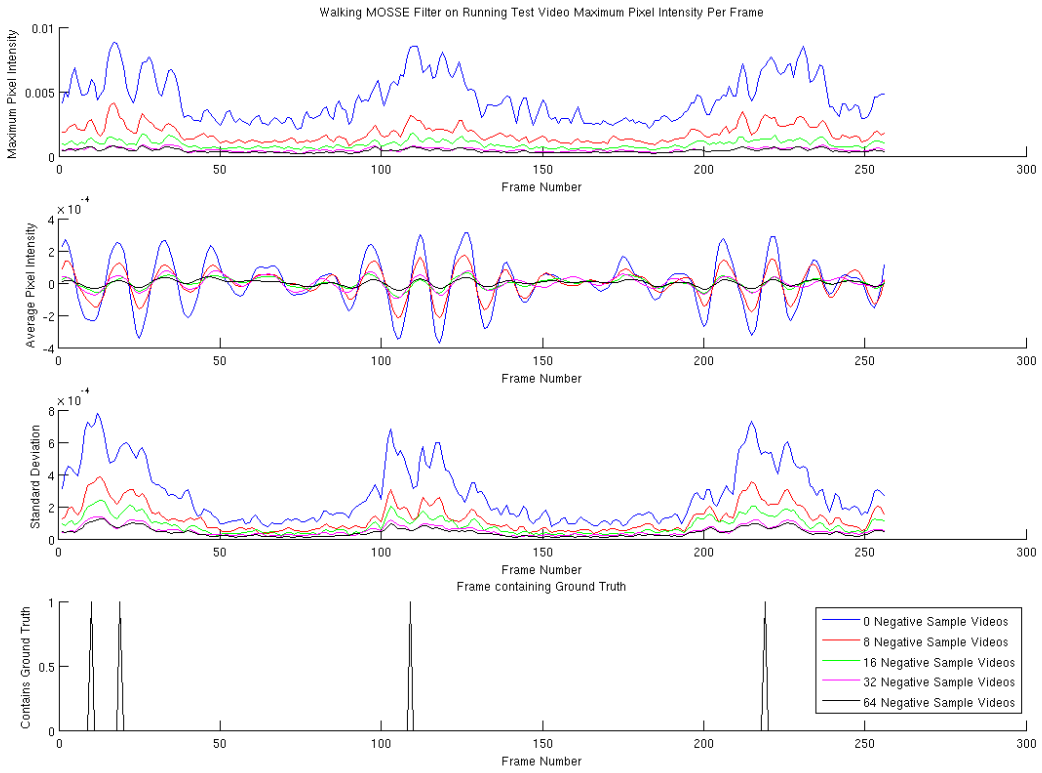

Figure 5.4: Statistics on walking MOSSE filter applied to running subject.

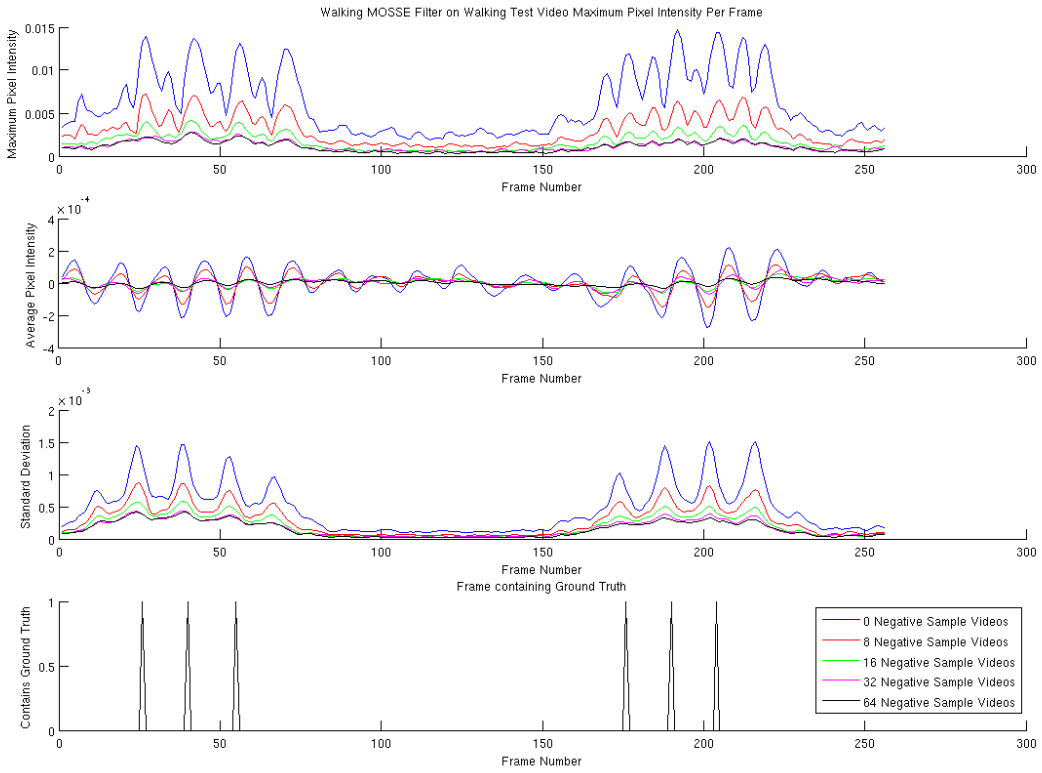

Figure 5.5: Statistics on walking MOSSE filter applied to walking subject.

negative samples. Since these 2 figures were created by using a MOSSE filter built with positive walking videos and then tested on handclapping and running videos, respectively, this is a very desirable effect.

However, notice that if we include too many negative training videos, the structure within the walking MOSSE filter tested on a walking video decreases greatly. Once we include 32 or 64 negative training videos, there is barely any difference among the 3 figures. Clearly, the choice of the number of negative training samples is very important. With no negative samples included in the walking MOSSE filter, the other actions appear as though they could have structure - there are several strong peaks in relation to other frames within the tested videos.

Given there is a large difference between the actions of {boxing, handclapping, handwaving} and {jogging, running, walking} we also provide similar examples in Figures 5.6 - 5.8. For these plots, the setup and experiment was similar. The difference is the MOSSE filter was trained using positive samples of handwaving videos instead of walking videos.

We notice the same effect, although to a lesser degree, in this set of examples as well. When the MOSSE filter was tested on the same action it was trained for, the inclusion of 8 negative samples did not kill the structure of the maximum pixel intensity and standard deviation. However, when the filter was tested on other actions, the apparent structure was quickly diminished. And once again, as we include too many negative training signals, all the structure in all figures (including the action the filter was trained to detect) was lost.

5.3 Three-Dimensional MOSSE Filters

In Chapter 3.3 we saw MOSSE filters created for each of the 4 target features from the FERET data set. In all examples, the MOSSE filter, when viewed in pixel space by

$$m^* = \mathcal{F}^{-1}\left(M^*\right)$$

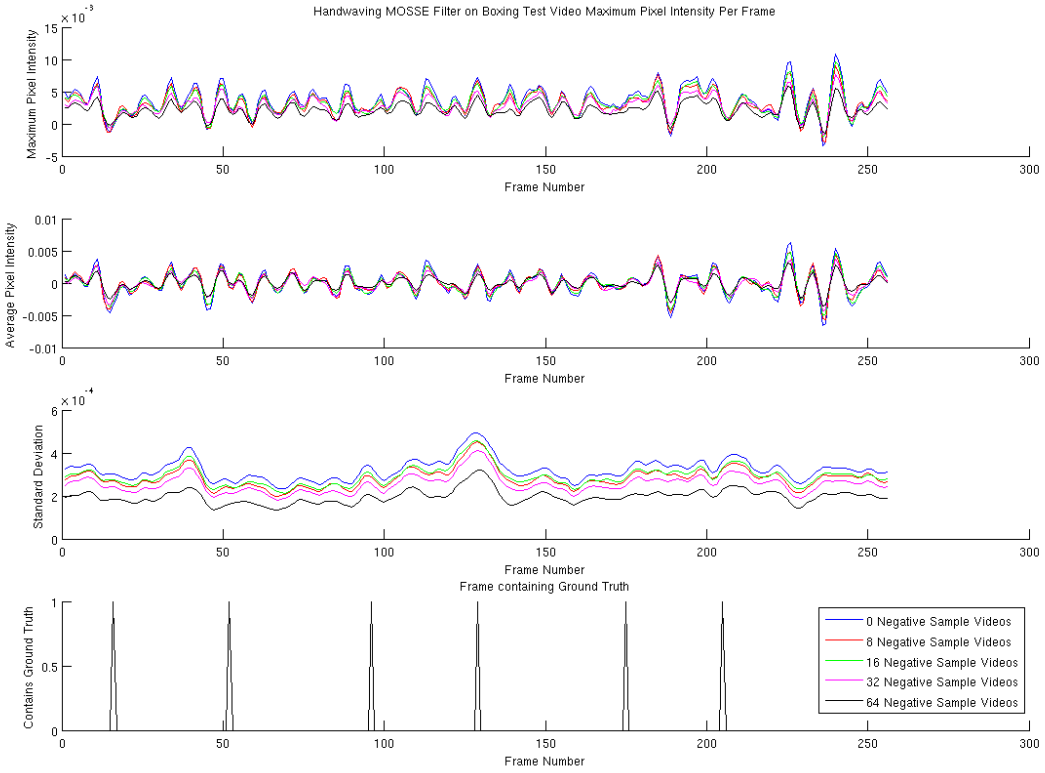

Figure 5.6: Statistics on handwaving MOSSE filter applied to boxing subject.

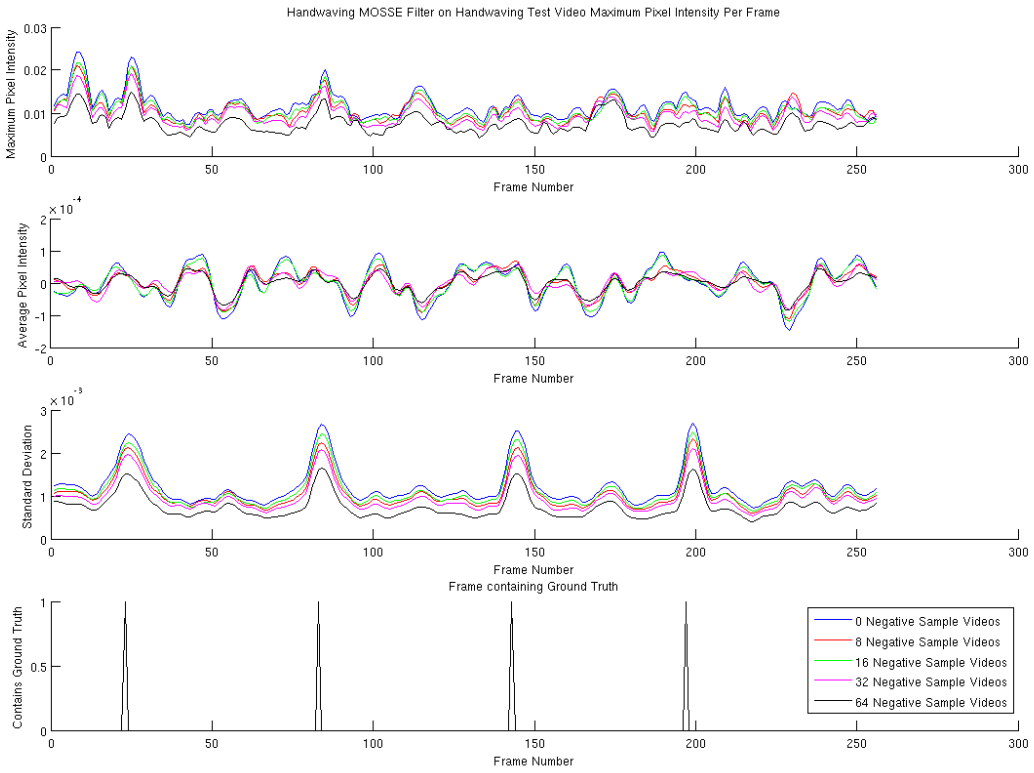

Figure 5.7: Statistics on handwaving MOSSE filter applied to handwaving subject.

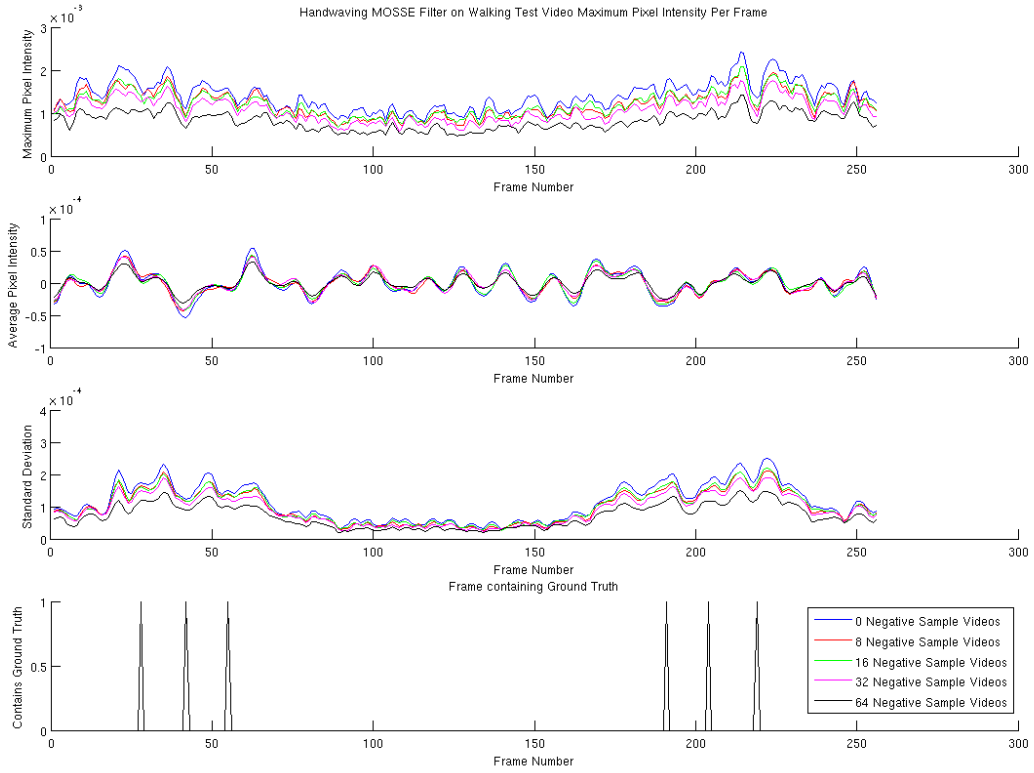

Figure 5.8: Statistics on handwaving MOSSE filter applied to walking subject.

placed the target feature in the center of the frame. In the same manner, the three-dimensional MOSSE filter operates the same way. As we will see in Figures 5.9 - 5.14, the main structure of the filter surrounds frame 128 and is centered in the frame. Now, we have to be careful in the construction of the filter. With the 2-D FERET data, there was only one example of a right eye or nose in each image. However, in the KTH data set we have multiple instances of a person boxing, or handwaving, per video. By looking at the ground truth coordinates in Appendix A we can see that these actions do not occur on the same frame across different videos. Yet the MOSSE filter only has structure in the action in the center of the pixel space MOSSE video.

For example, looking at selected frames of the boxing MOSSE filter in Figure 5.9, we see that near the front and end of the video (in terms of frames) there is very little structure. In any given original video featuring the action of boxing, we see that the action occurs about once every 30 frames (on average). However, the only structure within the MOSSE cube is located around frame 128, or the center of the data cube. This is because in the construction of the filter each instance of the desired action gets mapped to the center of the pixel MOSSE cube. Outside of a few frames in the center of m^* the rest of the m^* contains an average of everything that is not the action. This means that the description of the action is only a few frames in length and by applying the entire MOSSE filter to a test video we will only be adding noise to the response by including the non-action portion of the MOSSE filter.

5.4 Multi-Sigma 3-D Geometric Mean

We once again turn our attention to the Multi-Sigma Geometric Mean approach introduced in Section 4.8. The idea is that by using several σ values we will be able to separate true peaks, or detections, from false peaks with higher accuracy. Since our tests to reveal optimal values for a spatial and temporal σ in Chapter 5.1 provided inconclusive results for the optimal value of σ to use for each action, we will return to using spherical Gaussians. We hope that by selecting various sizes of Gaussians, the pixel response generated by applying

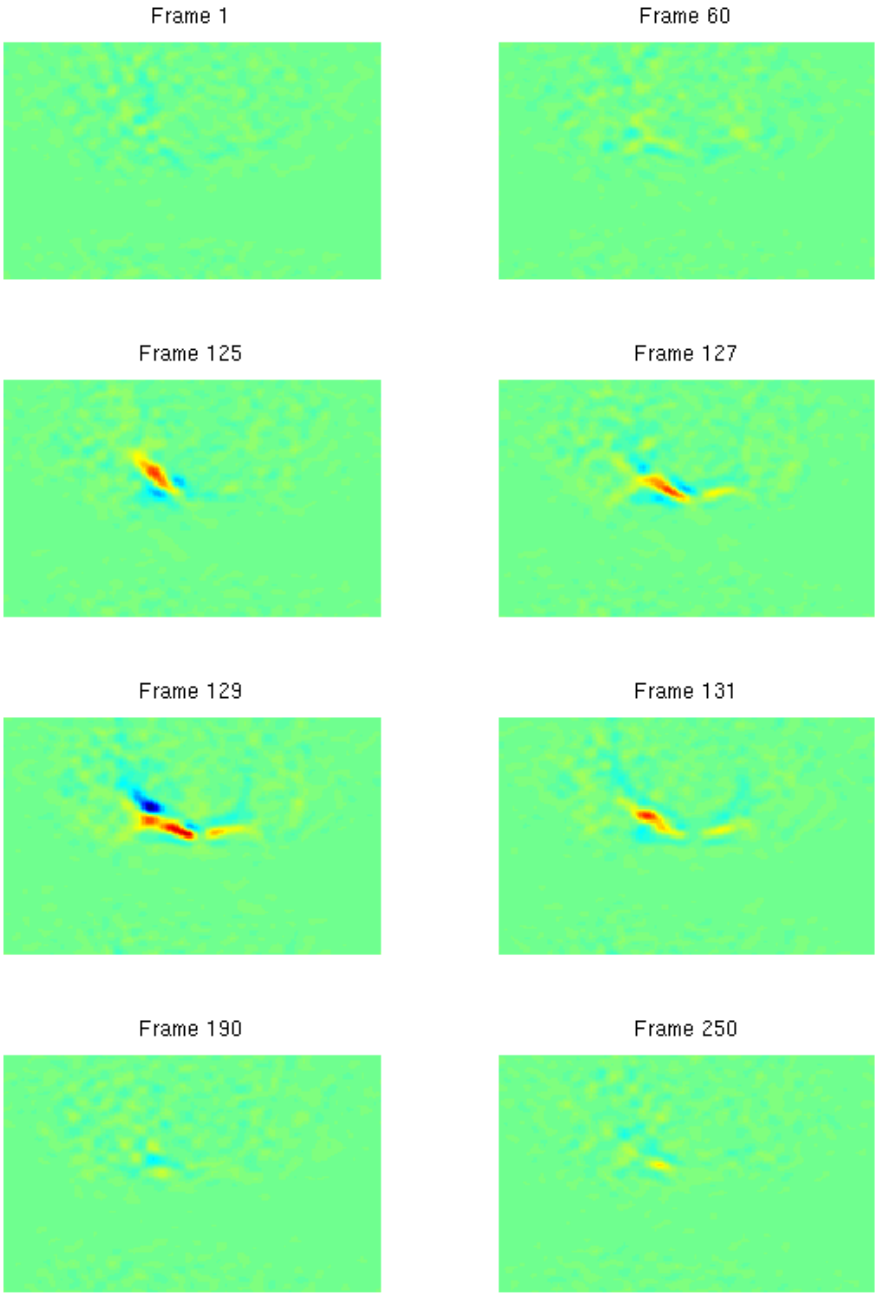

Figure 5.9: Selected frames from 3-D boxing MOSSE filter.

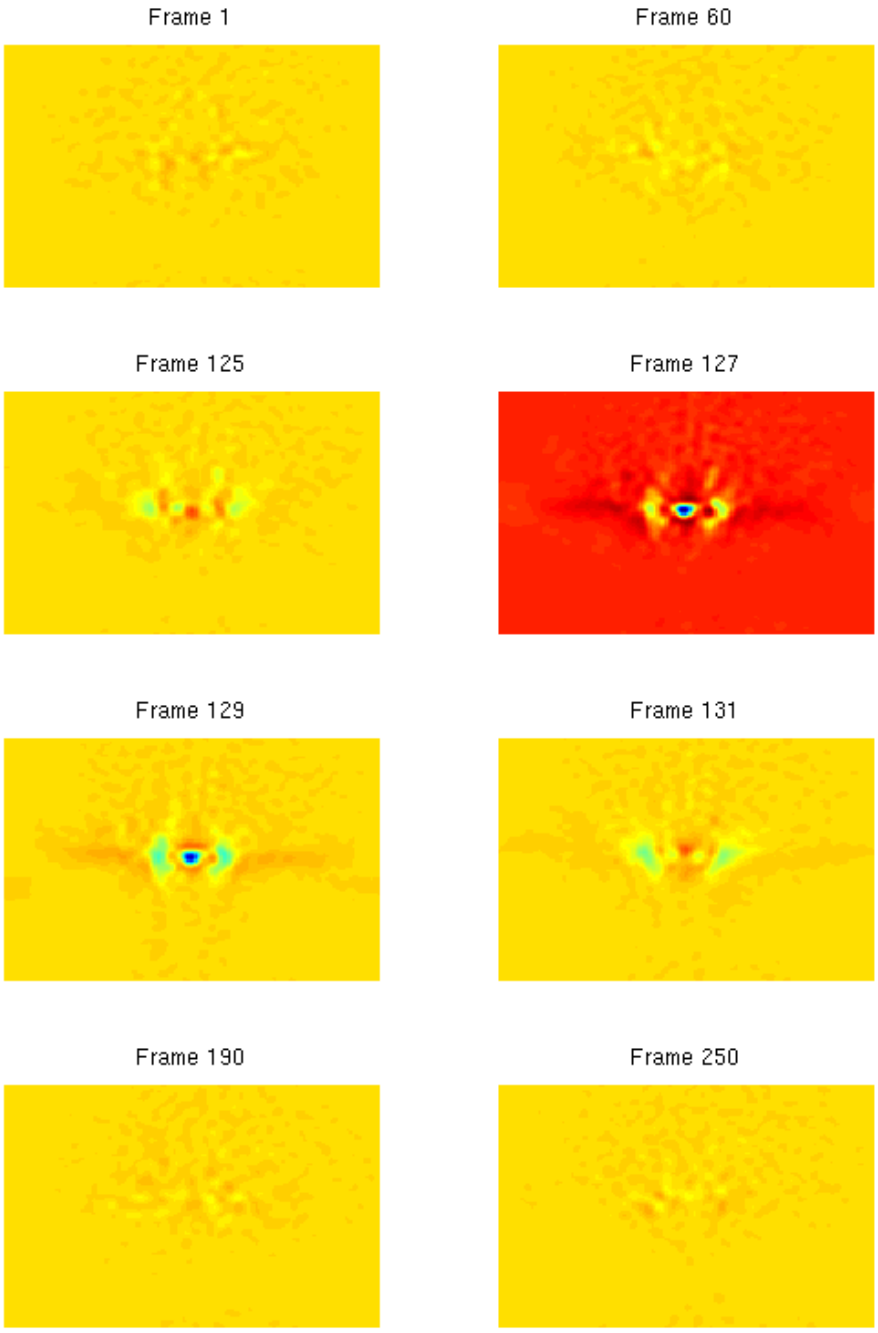

Figure 5.10: Selected frames from 3-D handclapping MOSSE filter.

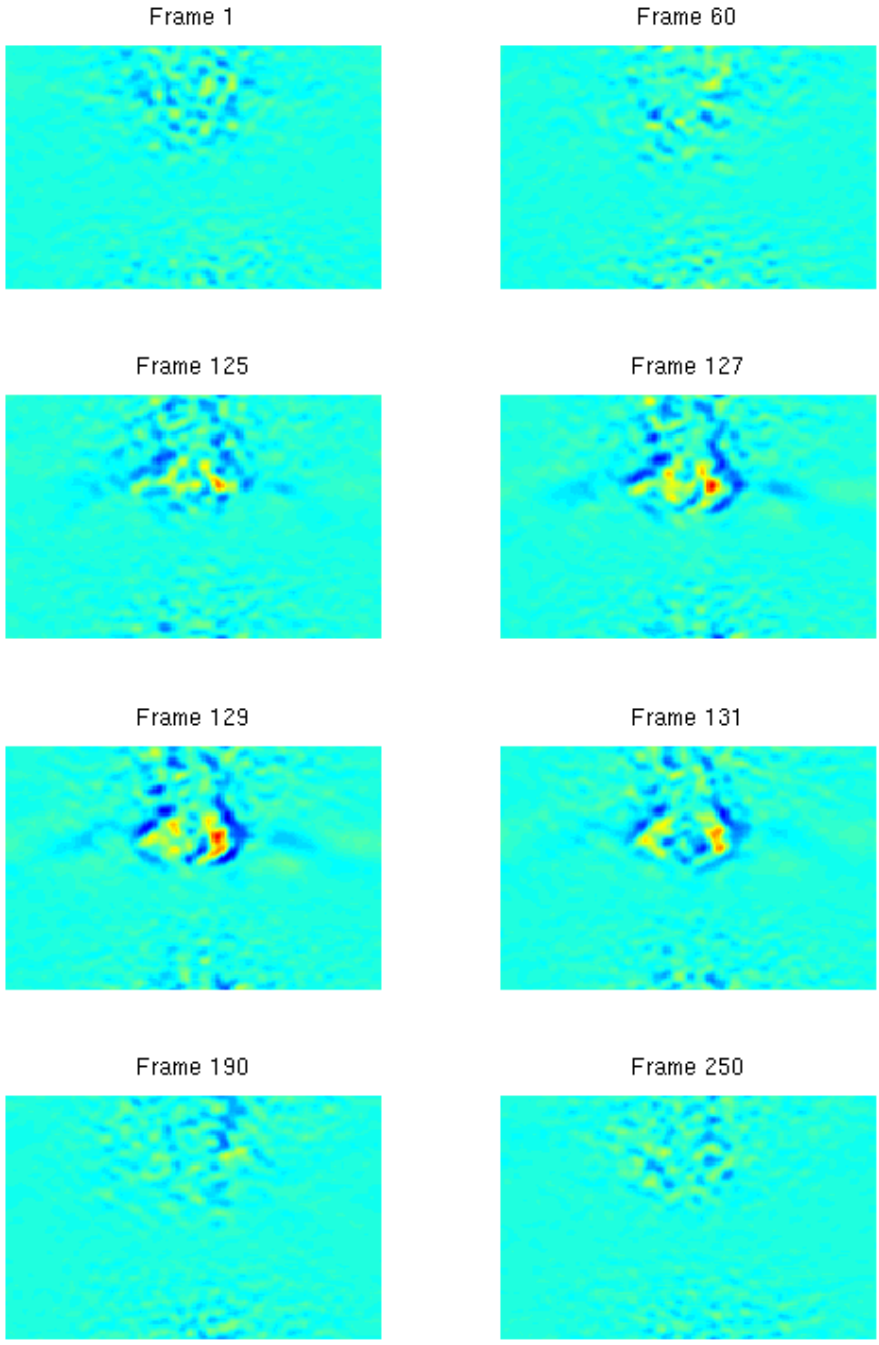

Figure 5.11: Selected frames from 3-D handwaving MOSSE filter.

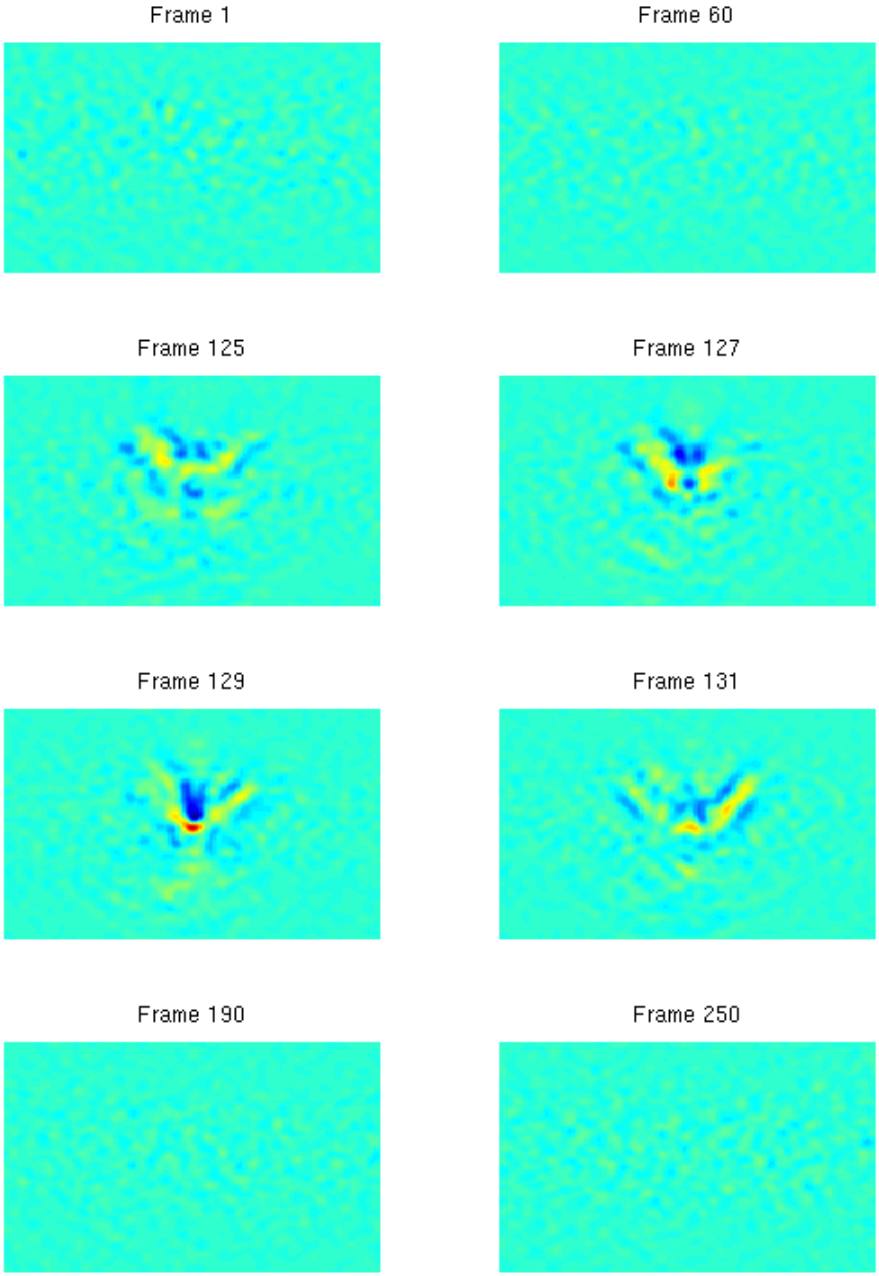

Figure 5.12: Selected frames from 3-D jogging MOSSE filter.

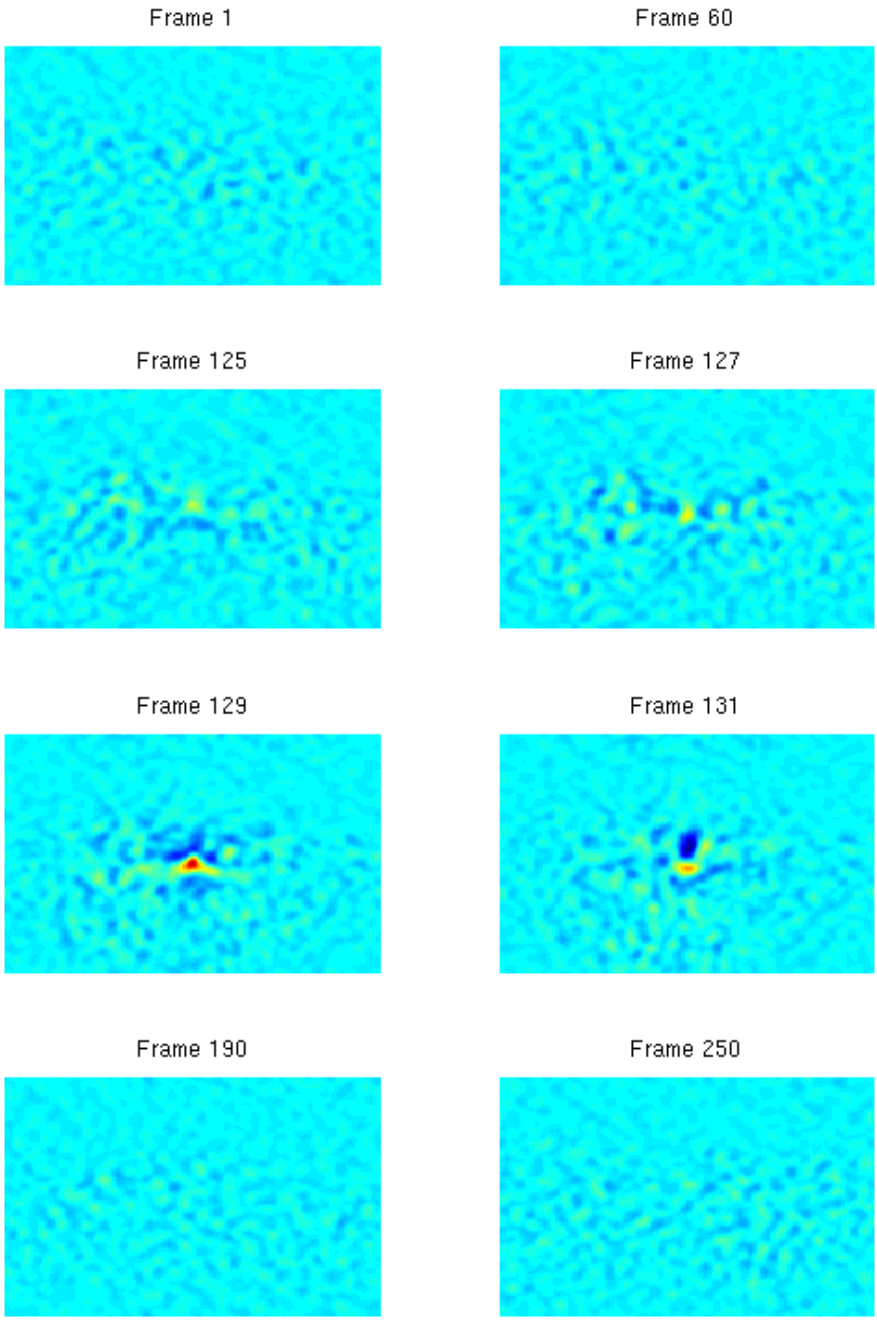

Figure 5.13: Selected frames from 3-D running MOSSE filter.

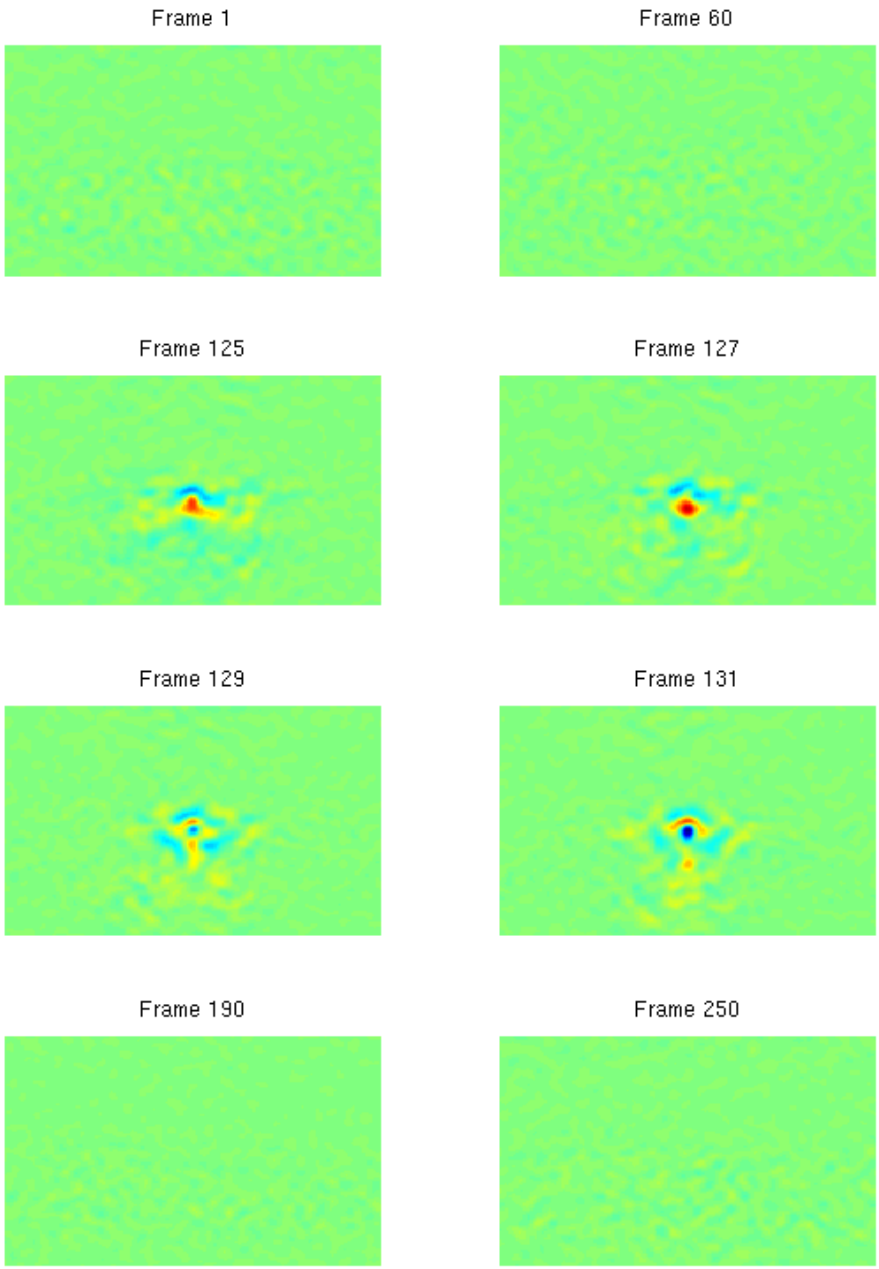

Figure 5.14: Selected frames from 3-D walking MOSSE filter.

a MOSSE filter to a test video will have different intensity values at locations where false peaks occur, and similar intensity values at locations where there is a true peak.

We provide two example cases in Figures 5.15 and 5.16. In the first case, we construct a MOSSE filter from a set of 96 handwaving videos. We then apply the 3-D MOSSE fitler to one of the remaining 4 handwaving videos that was not used in the construction of the MOSSE filter. We show the maximum pixel intensity and standard deviation per frame for σ values of 2, 4, and 8. Given that r_2 refers to the pixel response from applying a MOSSE filter made with a Gaussian with parameter $\sigma = 2$ with a test video, we have then computed

$$r_{2,4} = r_2 \odot r_4$$

and

$$r_{2,4,8} = r_{2,4} \odot r_8$$

In both figures, the red line provides a marking for which frame has been labeled as ground truth. By first looking at Figure 5.15 we notice there is a substantial improvement between the plots of r_2 and $r_{2,4}$. In $r_{2,4}$ the peaks are much more pronounced and it is easy to identify the true detections from the false detections.

However, this method is not entirely perfect. As we look at Figure 5.16 we see there is little structure to the plot of the maximum intensity value for $\sigma = 2$ when a MOSSE filter is created on the action of handwaving and tested on a video of walking. However, as we increase the value of σ and when we compute $r_{2,4}$ and $r_{2,4,8}$ we artificially generate very strong false detections when the person is walking through the video.

This initial results look encouraging. They illustrate that even with the too carefully chosen ground truth locations, other values of σ besides $\sigma = 1$ can be used to detect actions. If our conclusions concerning the 2-D Multi-Sigma Geometric Means hold, i.e., selecting values of σ to use in this product that are surrounding the global minimum of sum squared

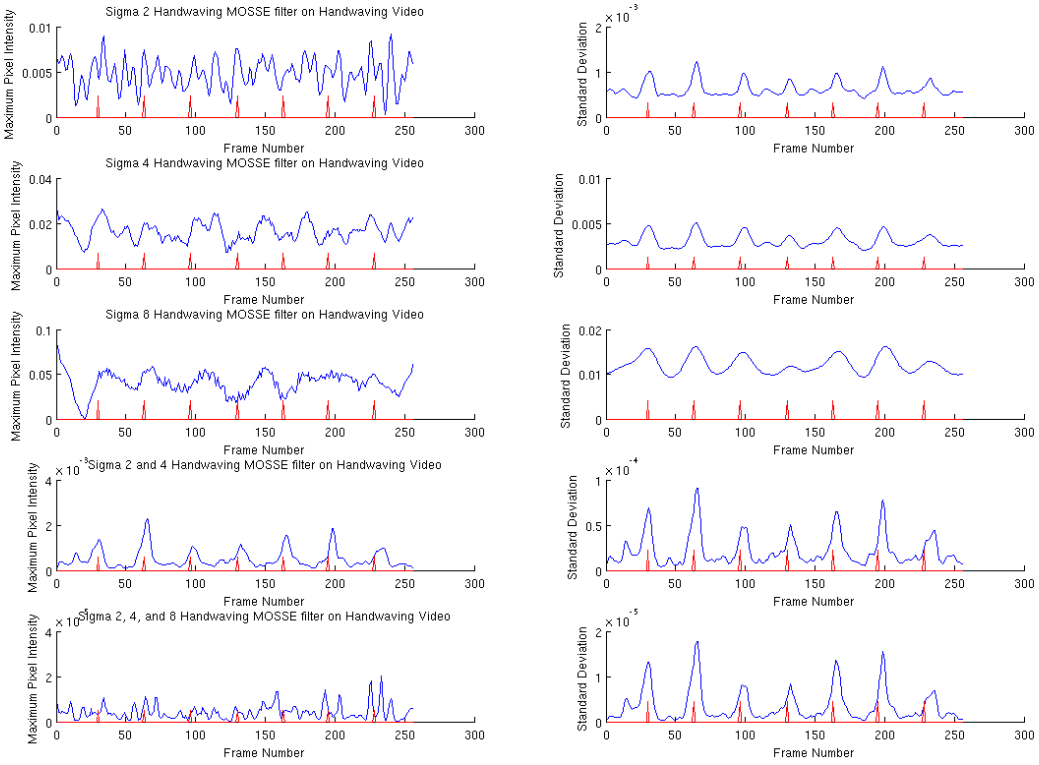

Figure 5.15: Example of using Multi-Sigma Geometric Mean on similar actions.

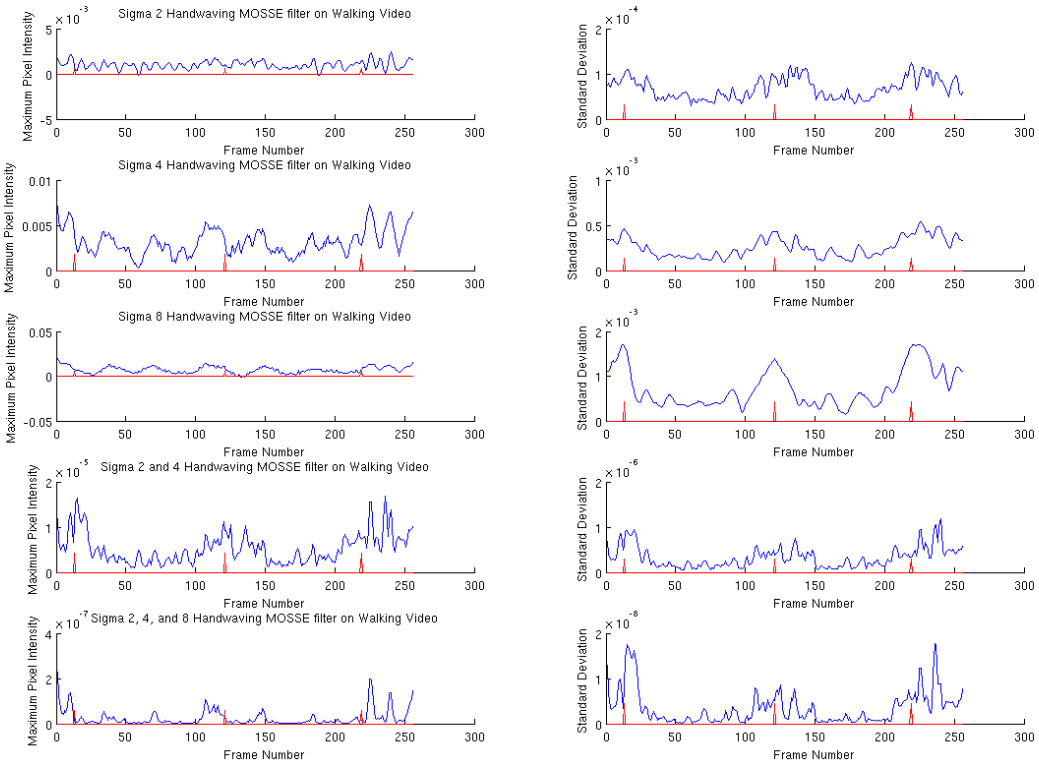

Figure 5.16: Example of using Multi-Sigma Geomtric Mean on dissimilar actions.

error will result in a higher accuracy of detection, then we can start searching for a range of σ values to use in this 3-D case.

5.5 KTH Conclusions

The KTH data set has been a widely investigated computer vision data set. Numerous publications have used KTH for action classification tests [5, 8, 15, 11, 14]. However, there are no published results concerning the task of action detection has (to the knowledge of this author). The results presented here suggest this remains a challenging problem.

In this chapter, we have extended the domain of application of MOSSE filters to 3-dimensional data i.e., video data cubes. We have developed several methods designed to improve the task of action detection. We have also explored some of the challenges that go along with action detection, such as selecting a threshold value to declare a detection has occurred and selecting a method for obtaining a value to be used in the thresholding. We have observed that the PSLR calculation is too computationally expensive to use in real time 3-D applications.

The KTH data set is possibly not the ideal data set to use in order to achieve a good understanding of MOSSE filters in 3-D. Given the low resolution, the small number of training samples, the difference between the 4 scenes varies to an extent that the filter has difficulties picking humans out of the scene, and various camera movements, shifts, zooms, and in-video camera focus changes are features which the MOSSE filter is not currently designed to handle.

Chapter 6

FUTURE WORK

There are several outstanding questions remaining concerning the MOSSE filter. Answering these questions would not only lead to a better understanding of the creation and application of the filter, but could lead to faster performance or increased accuracy of object/action classification and detection.

- How do the features present in the pixel space representation of the MOSSE filter correspond to the shape of the pixel response?
- What causes the difference in magnitude of different pixel responses despite the testing videos having been normalized?
- What information could we obtain if we were to apply the MOSSE filter over several features?
- How can we modify MOSSE filters to correct for scale, rotation, or temporal differences of an object/action?
- Depending on the data set, would it be advantageous to try rotations on the non-spherical Gaussian when creating the MOSSE filter?
- Can we determine a method of picking values of σ to use in the Multi-Sigma Geometric Mean method that will increase the intensity of true peaks in relation to false peaks?
- With three-dimensional data, how can we detect actions in the pixel response without knowing how many (if any) instances of the action occurs? Similarly in the two-dimensional case, what if there is more that one of the desired object present or what if the object is not present?

- Can we gain a better understanding of the filter by looking at its properties from a theoretical view-point through the use of Fourier analysis and other signal processing techniques?

- What data sets can we look at which are harder to perform detections on in 2-D and less difficult to perform detections on in 3-D?

In conclusion, it has been observed that the MOSSE filter is a powerful tool for object detection in images and action detection in video. The extension of the filter to the 3-D setting appears to have significant potential though more work needs to be done to realize the automatic surveillance system described in the introduction.

Bibliography

[1] David S. Bolme, *Theory and applications of optimized correlation output filters*, Ph.D. thesis, Colorado State University, 2011.

[2] David S. Bolme, J. Ross Beveridge, Bruce A. Draper, and Yi Man Lui, *Visual object tracking using adaptive correlation filters*, Computer Vision and Pattern Recognition (June 2010), 2544–2550.

[3] David S. Bolme, Bruce A. Draper, and J. Ross Beveridge, *Average of synthetic exact filters*, Computer Vision and Pattern Recognition (June 2009), 2105–2112.

[4] David S. Bolme, Yui Man Lui, Bruce A. Draper, and J. Ross Beveridge, *Simple real-time human detection using a single correlation filter*, PETS (2009), 1–8.

[5] M. Bregonzio, J. Li, S. Gong, and T. Xiang, *Discriminative topics modelling for action feature selection and recognition*, Proc. British Machine Vision Conference (BMVC) (2010), 1–11.

[6] B.V.K Kumar, A. Mahalanobis, S. Song, S.R.F. Sims, and J.F. Epperson, *Minimum squared error synthetic discriminant functions*, Optical Engineering **31** (1992), 915–922.

[7] Ivan Laptev and Barbara Caputo, *Kth website*, http://www.nada.kth.se/cvap/actions/, 2003.

[8] Y. M. Lui and J. R. Beveridge, *Tangent bundle for human action recognition*, Proc. IEEE Conference on Automatic Face and Gesture Recognition (FG2011) (2011), 97–102.

[9] A. Mahalanobis, BVK Vijaya Kumar, S. Song, SRF Sims, and JF Epperson, *Unconstrained correlation filters*, Applied Optics **33(17)** (1994), 3751–3759.

[10] David G. Messerschmitt, *Stationary points of a real-valued function of a complex variable*, Tech. Report UCB/EECS-2006-93, EECS Department, University of California, Berkeley, Jun 2006.

[11] S. O'Hara and B. A. Draper, *Scalable action recognition with a subspace forest*, Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) (2012), 1210–1217.

[12] P.J. Phillips, H.J. Moon, S.A. Rizvi, and P.J. Rauss, *The feret evaluation methodology for face-recognition algorithms*, IEEE Transactions of Pattern Analysis and Machine Intelligence **22(10)** (2000), 1090–1104.

[13] W. Press, B. Flanery, S. Teukolsky, and W. Vetterling, *Numerical recipes*, Cambridge Univ. Press, 1988.

[14] Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah, *Action mach a spatio-temporal maximum average correlation height filter for action recognition*, Computer Vision and Pattern Recognition (2008), 1–8.

[15] X. Wu, D. Xu, L. Duan, and J. Luo, *Action recognition using context and appearance distribution features*, Proc. Computer Vision and Pattern Recognition (CVPR) (2011), 489–496.

Appendix A

KTH GROUND TRUTH

Since there is no known ground truth data for the KTH data set, we provide our own, hand-labeled, ground truth data. All simulations, experiments, and trials in Chapter 5 were performed using the data provided here.

For each action we identified a spatio-temporal coordinate that we decided best represents the given action. To help illustrate the point chosen, we have placed a red dot on the location we have chosen in the following 6 figures. In boxing videos, we chose the furthest extended point on the top of the right hand as the arm was at its maximum extension during the action (see Figure A.1).

The handclapping spatio-temporal point is identified as the center of the top of the hands during the frame where the hands first make contact with each other see Figure A.2).

The handwaving spatio-temporal point is identified as the center of the top of the hands during the frame where the hands first overlap with each other. In actions where the hands do not overlap, the point is identified as the top of the right hand during the frame where the hands are the closest together (see Figure A.3).

The jogging spatio-temporal point is identified as the point where the two legs of a person intersect at the time when the angle of that point is maximized (see Figure A.4).

The running spatio-temporal point is identified as the point at the tip of the back foot (the foot that is not on the ground) when the back leg is parallel to the ground. In cases where the leg does not reach that state, we take the tip of the back foot when the leg is as close to parallel to the ground as possible (see Figure A.5).

The walking spatio-temporal point is identified as the point of the tip of the back foot as soon as the front foot is firmly on the ground (see Figure A.6).

As with any human-labeling of data, there will be errors in ground truth data. For practical applications it is recommended that a variance of 10 pixel in the spatial dimension

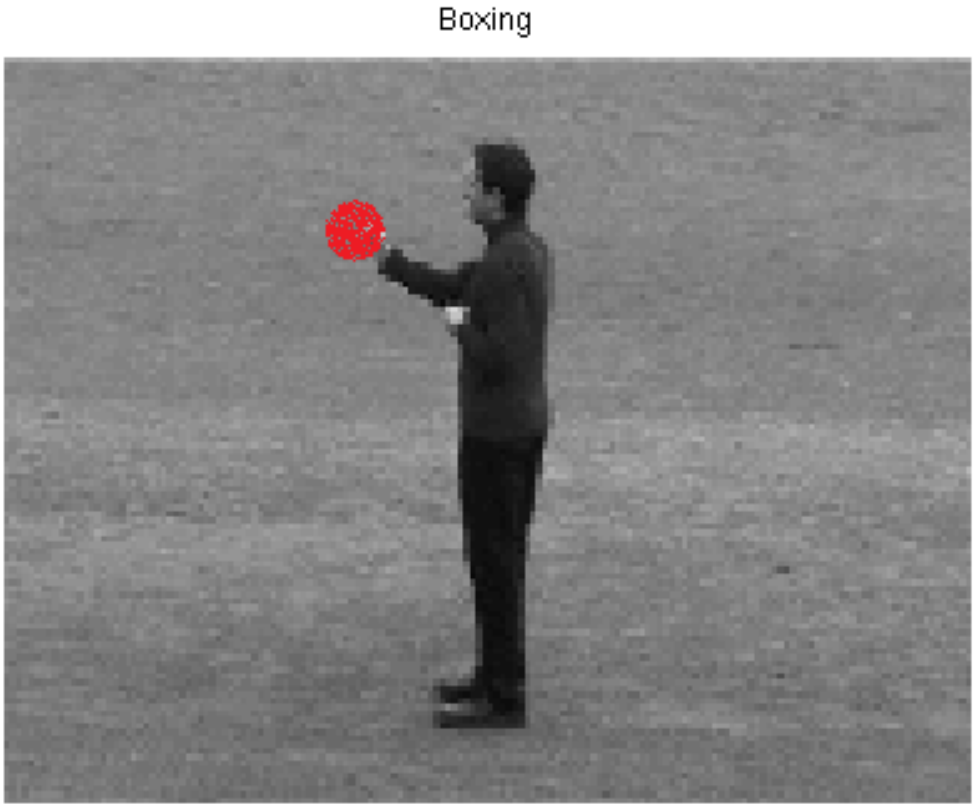

Figure A.1: Boxing labeling example.

Figure A.2: Handclapping labeling example.

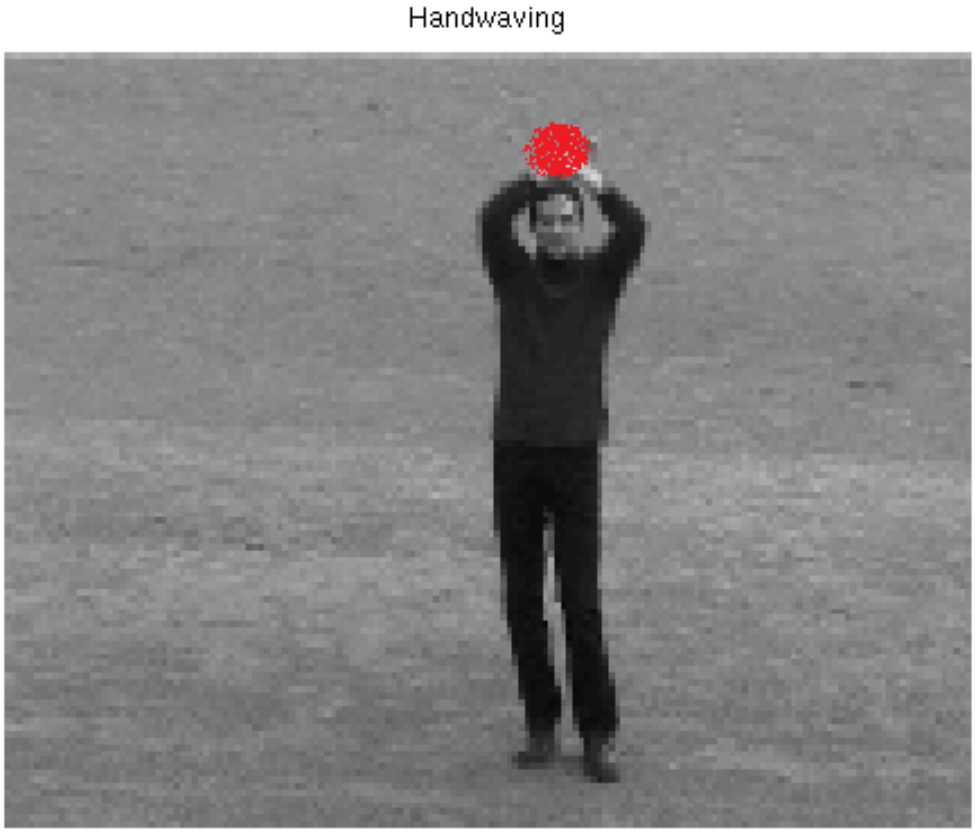

Figure A.3: Handwaving labeling example.

Figure A.4: Jogging labeling example.

Figure A.5: Running labeling example.

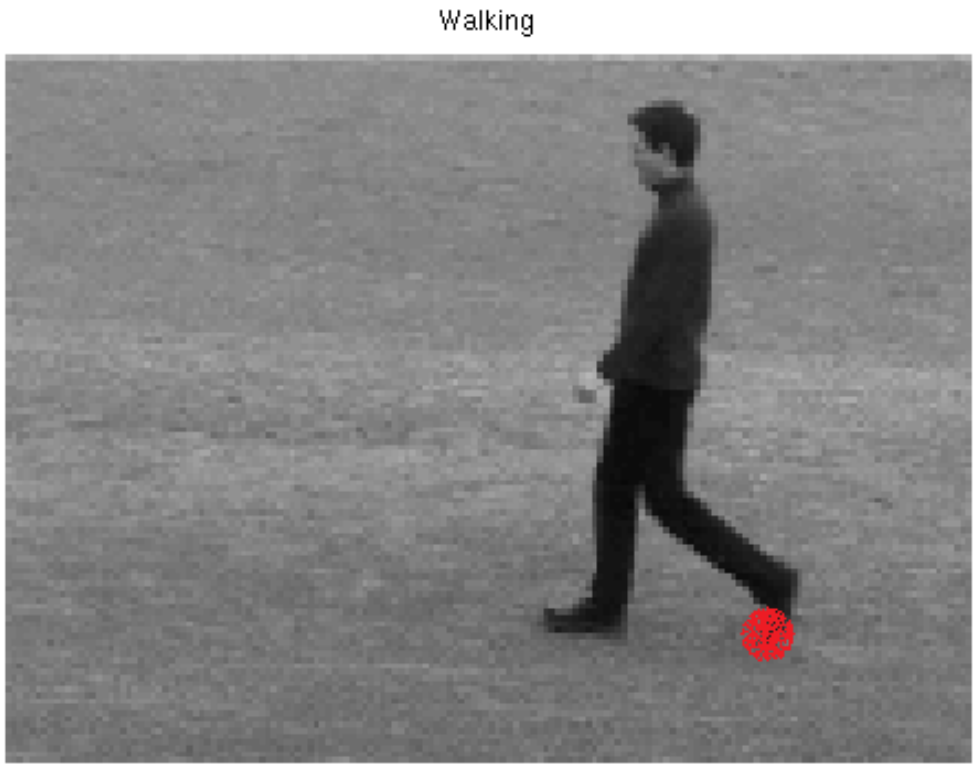

Figure A.6: Walking labeling example.

and 2 pixels (or frames) in the temporal dimension (when measured using the l_∞-norm be used. Tables A.7 - A.26 provide all ground truth data used. It is separated by Person Number and Scene Number as provided with the download of the original KTH data [7]. The column denoted as y gives the column number. The column denoted as x gives the row number. And the column denoted as t gives the frame number where the first column/row/frame is numbered as 1.

Person #	Scene #	y	x	t
1	1	33	58	11
	1	31	58	25
	1	31	56	40
	1	32	60	53
	1	31	59	67
	1	30	59	81
	1	29	58	95
	1	29	58	109
	1	29	57	123
	1	32	58	137
	1	31	56	151
	1	31	55	165
	1	31	56	180
	1	33	58	194
	1	32	58	209
	1	32	58	223
	1	32	58	238
	2	31	77	10
	2	26	68	24
	2	28	74	38
	2	33	75	52
	2	39	77	67
	2	39	77	80
	2	37	71	96
	2	33	68	111
	2	33	64	125
	2	31	67	140
	2	32	69	154
	2	36	74	170
	2	37	77	184
	2	35	75	198
	2	31	67	214
	2	29	65	227
	2	33	67	242
	3	27	60	12
	3	31	63	26
	3	29	63	39
	3	32	62	70
	3	30	61	85
	3	31	63	101
	3	32	63	117
	3	30	63	133
	3	30	64	148
	3	31	64	163
	3	30	63	178
	3	32	65	194
	3	31	63	210
	3	31	63	225
	3	31	61	239
	4	28	51	7
	4	29	49	27
	4	28	48	46
	4	25	49	61
	4	27	49	76
	4	26	47	90
	4	27	46	104
	4	25	50	118
	4	25	48	131
	4	25	50	146
	4	25	54	159
	4	27	50	174
	4	28	49	188
	4	26	48	201
	4	25	47	215
	4	26	48	229
	4	25	49	242
10	1	21	63	6
	1	20	67	37
	1	23	65	67
	1	24	65	100
	1	23	67	131
	1	23	67	164
	1	23	66	196
	1	22	65	210
	1	24	66	227
	2	26	62	7
	2	22	62	36
	2	19	68	68
	2	21	62	102
	2	26	69	135

Person #	Scene #	y	x	t
	2	26	73	170
	2	23	70	205
	2	19	63	242
	3	19	51	9
	3	19	49	41
	3	20	49	73
	3	18	52	108
	3	21	52	143
	3	20	51	180
	3	19	50	215
	4	19	50	7
	4	17	48	38
	4	17	48	71
	4	19	50	105
	4	21	48	138
	4	20	49	172
	4	21	48	205
	4	19	49	239
11	1	26	46	19
	1	29	43	59
	1	30	47	108
	1	32	48	147
	1	29	47	191
	1	31	44	228
	2	27	58	17
	2	33	63	83
	2	40	71	120
	2	31	71	172
	2	31	68	233
	3	30	56	23
	3	31	55	69
	3	27	57	116
	3	28	59	188
	3	27	59	242
	4	19	45	11
	4	16	51	20
	4	19	46	81
	4	21	45	91
	4	14	51	174
	4	13	55	234
12	1	28	44	60
	1	29	39	129
	1	29	41	158
	1	29	38	211
	2	35	69	16
	2	34	74	52
	2	43	77	96
	2	29	70	129
	2	44	80	175
	2	25	77	205
	3	22	60	42
	3	23	62	80
	3	26	59	130
	3	30	59	196
	3	30	59	228
	4	22	43	22
	4	19	37	77
	4	20	28	183
	4	22	29	205
13	1	21	55	19
	1	25	55	40
	1	25	57	63
	1	27	57	82
	1	25	59	102
	1	25	60	122
	1	32	59	142
	1	32	57	162
	1	29	58	184
	1	31	59	202
	1	32	58	223
	1	29	59	244
	2	33	94	12
	2	29	87	33
	2	27	88	56
	2	40	89	77
	2	34	92	98
	2	23	91	120
	2	41	88	141
	2	44	89	162
	2	38	90	182

Person #	Scene #	y	x	t
	2	31	86	202
	2	33	87	222
	2	40	87	242
	3	31	75	21
	3	33	76	43
	3	34	75	64
	3	33	74	86
	3	32	75	105
	3	35	78	127
	3	35	76	149
	3	33	78	209
	3	33	78	230
	4	22	52	11
	4	22	47	33
	4	24	46	54
	4	25	46	74
	4	25	49	96
	4	25	48	115
	4	29	48	139
	4	29	50	160
	4	28	49	180
	4	32	52	202
	4	29	52	223
	4	25	51	246
14	1	29	109	13
	1	29	111	52
	1	34	112	93
	1	29	111	133
	1	31	114	172
	1	29	113	211
	1	29	117	248
	2	32	110	38
	2	27	120	77
	2	44	102	118
	2	38	104	158
	2	30	115	195
	2	42	105	231
	3	29	106	10
	3	31	107	51
	3	33	105	96
	3	31	106	132
	3	36	108	175
	3	34	107	214
	4	19	81	27
	4	20	81	67
	4	25	84	103
	4	25	84	140
	4	18	83	179
	4	18	83	217
15	1	29	59	10
	1	31	57	45
	1	32	57	80
	1	36	57	122
	1	34	56	164
	1	36	56	204
	1	36	54	244
	2	24	56	25
	2	33	70	65
	2	26	68	105
	2	27	65	142
	2	38	73	184
	2	38	71	223
	3	29	64	27
	3	27	67	64
	3	27	68	104
	3	28	70	142
	3	29	69	181
	3	34	71	218
	4	29	38	24
	4	31	37	61
	4	31	37	100
	4	31	37	137
	4	32	39	172
	4	29	37	207
	4	31	35	244
16	1	33	62	14
	1	34	58	38
	1	32	61	48
	1	33	62	71
	1	37	64	96

Figure A.7: Boxing Coordinates.

Person #	Scene #	y	x	t
	1	34	64	116
	1	36	63	124
	1	35	64	146
	1	35	62	169
	1	36	61	193
	1	36	60	215
	1	38	62	237
	2	17	84	13
	2	12	86	26
	2	37	85	51
	2	37	87	75
	2	34	90	98
	2	29	93	124
	2	25	97	148
	2	37	91	170
	2	38	94	188
	2	35	97	211
	2	34	99	221
	2	31	102	242
	3	29	30	16
	3	28	36	42
	3	31	33	66
	3	31	36	92
	3	31	36	107
	3	37	38	132
	3	35	44	142
	3	34	43	142
	3	34	45	190
	3	37	51	215
	3	37	49	239
	4	23	63	21
	4	19	63	48
	4	19	62	104
	4	19	64	164
	4	22	68	226
17	1	28	58	21
	1	25	60	47
	1	30	60	76
	1	26	61	103
	1	28	63	129
	1	26	64	156
	1	27	64	185
	1	27	66	210
	1	25	65	237
	2	21	71	25
	2	20	70	57
	2	26	75	88
	2	21	70	124
	2	26	74	157
	2	22	70	188
	2	25	74	219
	2	31	74	248
	3	29	57	30
	3	26	57	64
	3	26	58	96
	3	22	55	131
	3	25	57	167
	3	27	57	200
	3	27	59	233
	4	15	56	6
	4	12	55	30
	4	12	56	52
	4	16	56	74
	4	11	58	98
	4	13	57	119
	4	14	59	141
	4	13	58	162
	4	14	57	184
	4	26	55	207
	4	23	55	227
	4	21	57	248
18	1	20	48	5
	1	20	49	48
	1	21	47	97
	1	20	47	147
	1	21	48	201
	1	21	52	249
	2	27	52	17
	2	30	58	69
	2	31	59	118

Person #	Scene #	y	x	t
	2	28	59	172
	2	35	62	231
	3	30	48	10
	3	28	48	58
	3	26	48	112
	3	27	47	165
	3	31	47	220
	4	27	73	7
	4	28	71	42
	4	26	67	80
	4	24	70	122
	4	24	74	168
	4	25	73	217
19	1	34	120	27
	1	33	122	42
	1	32	122	58
	1	30	121	86
	1	30	120	118
	1	32	120	147
	1	33	122	176
	1	28	120	205
	1	30	120	235
	2	32	67	8
	2	21	59	38
	2	20	63	70
	2	30	60	103
	2	34	66	134
	2	38	69	167
	2	42	71	197
	2	29	66	230
	3	27	54	38
	3	26	57	70
	3	27	52	102
	3	28	52	131
	3	30	54	163
	3	31	55	196
	3	27	53	225
	4	21	112	12
	4	23	115	43
	4	19	120	73
	4	22	119	103
	4	24	119	131
	4	22	123	161
	4	23	123	191
	4	25	123	217
	4	24	124	245
2	1	20	50	6
	1	19	50	33
	1	14	52	59
	1	14	53	85
	1	22	53	114
	1	16	55	143
	1	16	52	176
	1	24	54	205
	2	29	56	8
	2	49	72	38
	2	36	67	68
	2	28	55	101
	2	39	67	137
	2	26	56	173
	2	30	58	211
	2	36	63	244
	3	28	99	8
	3	29	104	40
	3	38	105	72
	3	31	109	101
	3	37	104	130
	3	28	108	163
	3	34	110	196
	3	32	109	228
	4	27	65	34
	4	14	63	71
	4	19	62	114
	4	22	61	157
	4	19	65	198
	4	21	63	240
20	1	21	102	23
	1	24	103	61
	1	26	106	102
	1	26	106	139

Person #	Scene #	y	x	t
	1	27	106	177
	1	23	107	213
	2	29	104	26
	2	25	108	57
	2	28	109	92
	2	34	101	125
	2	32	105	158
	2	24	110	190
	2	23	111	224
	3	23	80	22
	3	25	79	54
	3	23	84	86
	3	26	88	119
	3	21	88	151
	3	25	87	184
	3	27	90	216
	3	27	90	249
	4	19	99	7
	4	19	98	38
	4	17	98	69
	4	16	97	100
	4	16	97	131
	4	17	97	161
	4	17	97	192
	4	19	98	221
21	1	25	70	16
	1	17	55	60
	1	26	60	97
	1	17	55	119
	1	23	52	140
	1	22	61	164
	1	24	54	172
	1	22	55	191
	1	22	60	209
	1	18	50	226
	1	19	50	238
	2	25	71	14
	2	31	74	30
	2	21	70	53
	2	25	69	69
	2	26	71	83
	2	19	61	104
	2	22	64	119
	2	17	55	137
	2	17	48	149
	2	29	65	167
	2	31	62	186
	2	29	57	201
	2	26	60	217
	2	20	58	232
	2	18	56	245
	3	17	55	12
	3	19	55	34
	3	16	49	55
	3	24	50	82
	3	13	43	106
	3	17	43	125
	3	13	41	147
	3	10	42	169
	3	15	43	192
	3	13	39	204
	3	16	42	230
	3	14	49	249
	4	17	39	18
	4	13	36	37
	4	19	44	55
	4	14	36	82
	4	15	34	97
	4	11	30	117
	4	15	39	136
	4	17	35	153
	4	15	30	170
	4	21	34	186
	4	20	40	203
	4	19	35	227
	4	17	31	239
22	1	18	30	118
	1	22	31	185
	1	24	22	191
	1	18	23	231

Figure A.8: Boxing Coordinates Continued.

Person #	Scene #	y	x	t
	2	15	57	14
	2	19	51	20
	2	21	57	60
	2	9	36	112
	2	16	42	157
	2	18	41	163
	2	24	37	232
	3	19	63	13
	3	17	56	45
	3	20	41	51
	3	17	24	82
	3	17	17	136
	3	23	7	142
	3	20	10	200
	3	22	3	206
	3	20	1	238
	4	14	105	8
	4	16	112	13
	4	14	115	54
	4	17	122	109
	4	13	123	150
	4	14	113	216
	4	16	119	220
23	1	23	113	5
	1	19	113	43
	1	21	115	79
	1	25	112	150
	2	25	113	101
	2	18	120	110
	2	33	109	138
	2	31	103	166
	2	31	106	177
	3	29	101	55
	3	20	112	78
	3	24	107	93
	3	17	112	187
	3	16	127	203
	3	25	94	21
	4	29	83	36
	4	21	81	44
	4	26	81	86
	4	23	78	94
	4	22	84	215
24	1	27	113	19
	1	28	120	44
	1	29	121	65
	1	25	122	86
	1	24	121	107
	1	27	122	131
	1	26	123	152
	1	27	123	174
	1	25	124	196
	1	26	125	219
	1	26	124	242
	2	34	114	3
	2	32	113	24
	2	36	107	45
	2	39	102	67
	2	40	103	88
	2	27	122	110
	2	27	131	128
	2	29	122	150
	2	32	115	171
	2	35	113	190
	2	38	109	213
	2	39	105	232
	3	25	118	15
	3	28	118	36
	3	27	121	57
	3	25	121	79
	3	26	119	99
	3	26	120	121
	3	28	121	142
	3	28	119	162
	3	30	121	185
	3	29	122	204
	3	29	120	225
	3	29	122	245
	4	28	103	4
	4	32	107	23

Person #	Scene #	y	x	t
	4	30	108	43
	4	31	108	64
	4	29	107	84
	4	29	107	105
	4	29	109	125
	4	29	107	146
	4	31	110	165
	4	31	107	185
	4	30	108	205
	4	31	105	227
	4	33	108	242
25	1	23	56	27
	1	23	54	64
	1	20	53	102
	1	20	55	142
	1	22	56	185
	1	22	56	228
	2	20	61	10
	2	22	64	45
	2	27	65	81
	2	29	67	116
	2	25	65	153
	2	10	61	188
	2	31	71	224
	3	20	63	30
	3	21	61	74
	3	23	61	116
	3	23	63	160
	3	26	64	207
	4	24	56	7
	4	19	60	42
	4	22	60	78
	4	19	60	112
	4	20	62	145
	4	19	61	180
	4	21	60	215
	4	22	59	249
3	1	31	66	7
	1	33	70	33
	1	33	66	61
	1	27	68	88
	1	25	66	114
	1	22	67	142
	1	24	66	167
	1	24	67	194
	1	22	67	220
	1	21	67	247
	2	14	5	25
	2	32	67	57
	2	37	70	89
	2	35	71	120
	2	31	65	151
	2	27	64	185
	2	9	52	219
	3	29	54	22
	3	30	54	51
	3	28	57	82
	3	29	56	111
	3	26	61	142
	3	28	60	175
	3	23	63	203
	3	20	68	235
	4	17	51	25
	4	21	53	56
	4	17	53	90
	4	17	52	121
	4	21	54	151
	4	18	55	181
	4	19	56	214
	4	15	55	246
4	1	18	71	9
	1	19	71	63
	1	20	73	117
	1	20	73	168
	1	20	73	220
	2	22	63	9
	2	26	67	57
	2	27	64	102
	2	21	60	147
	2	28	65	196

Person #	Scene #	y	x	t
	2	25	63	239
	3	14	66	9
	3	14	67	54
	3	19	64	100
	3	17	68	144
	3	17	71	183
	3	16	73	226
	4	21	40	9
	4	20	39	54
	4	20	37	99
	4	21	38	140
	4	20	40	188
	4	21	39	233
5	1	28	51	20
	1	26	51	50
	1	26	52	80
	1	26	51	112
	1	27	51	145
	1	27	54	176
	1	24	53	210
	1	24	52	245
	2	29	55	6
	2	31	61	32
	2	28	56	58
	2	26	57	85
	2	35	66	112
	2	32	64	138
	2	20	52	166
	2	17	53	195
	2	24	56	223
	3	23	58	6
	3	25	60	29
	3	24	60	55
	3	24	59	79
	3	25	56	102
	3	27	56	130
	3	29	61	154
	3	26	60	181
	3	27	60	206
	3	28	61	231
	4	26	50	19
	4	21	49	45
	4	19	49	70
	4	21	49	95
	4	24	46	121
	4	23	46	147
	4	24	46	173
	4	21	44	200
	4	26	45	226
6	1	17	76	10
	1	30	74	31
	1	27	69	49
	1	28	70	59
	1	33	75	74
	1	27	77	96
	1	28	73	105
	1	35	78	138
	1	32	81	160
	1	31	82	183
	1	26	79	199
	1	31	76	206
	1	32	81	243
	2	33	70	8
	2	29	75	22
	2	27	71	52
	2	29	66	80
	2	35	70	109
	2	47	69	117
	2	40	65	134
	2	26	57	144
	2	22	65	205
	2	31	71	228
	3	32	76	16
	3	25	75	39
	3	28	68	68
	3	27	74	85
	3	29	72	112
	3	29	71	122
	3	29	71	149
	3	29	78	174

Figure A.9: Boxing Coordinates Continued.

Person #	Scene #	y	x	t
	3	33	71	198
	3	27	69	239
	4	32	68	23
	4	30	67	55
	4	27	66	66
	4	22	67	94
	4	21	67	106
	4	25	64	131
	4	26	66	154
	4	24	63	166
	4	27	66	192
	4	24	68	205
	4	28	68	235
7	1	25	48	5
	1	27	46	28
	1	32	52	51
	1	29	52	70
	1	28	48	91
	1	29	48	112
	1	32	49	132
	1	30	50	153
	1	28	51	172
	1	27	52	192
	1	27	52	212
	1	27	55	233
	2	30	61	4
	2	24	60	28
	2	30	65	55
	2	32	69	77
	2	32	68	102
	2	22	60	127
	2	18	64	150
	2	24	65	174
	2	30	66	198
	2	31	67	223
	3	26	67	25
	3	28	71	51
	3	29	77	75
	3	25	75	97
	3	27	77	121
	3	29	76	142
	3	30	79	164
	3	30	79	184
	3	31	80	205
	3	35	57	6
	4	36	60	36
	4	37	58	67
	4	38	60	95
	4	32	59	122
	4	33	58	152
	4	35	56	180
	4	37	56	209
8	1	28	54	20
	1	26	54	41
	1	26	56	68
	1	26	57	90
	1	28	60	112
	1	27	59	137
	1	25	57	163
	1	29	56	187
	1	27	53	212
	1	25	56	235
	2	24	49	19
	2	35	64	44
	2	20	40	70
	2	20	49	95
	2	17	51	122
	2	23	57	149
	2	28	59	175
	2	32	62	199
	2	35	62	226
	3	19	64	5
	3	18	64	27
	3	20	63	51
	3	21	61	76
	3	20	63	99
	3	19	60	125
	3	21	62	148
	3	22	63	173
	3	25	62	197

Person #	Scene #	y	x	t
	3	22	62	221
	3	23	62	246
	4	24	106	26
	4	24	106	52
	4	24	106	77
	4	21	107	103
	4	23	106	126
	4	24	106	150
	4	23	107	175
	4	24	108	199
	4	23	107	225
9	1	29	56	21
	1	28	57	42
	1	27	60	65
	1	26	60	90
	1	29	58	113
	1	27	56	138
	1	26	59	162
	1	29	57	187
	1	27	56	213
	1	27	57	234
	2	27	51	18
	2	35	66	43
	2	32	62	70
	2	23	44	95
	2	20	53	122
	2	23	58	148
	2	30	59	174
	2	33	62	199
	2	35	63	224
	2	39	68	249
	3	20	68	4
	3	19	66	29
	3	22	62	53
	3	23	63	75
	3	22	65	99
	3	23	60	124
	3	22	63	148
	3	23	64	172
	3	26	62	197
	3	24	64	221
	3	24	64	245
	4	28	107	5
	4	26	104	26
	4	27	105	51
	4	25	105	75
	4	23	106	101
	4	25	105	125
	4	25	104	149
	4	25	104	151
	4	25	104	176
	4	24	104	198
	4	24	105	224

Figure A.10: Boxing Coordinates Continued.

Person #	Scene #	y	x	t
1	1	37	93	9
	1	42	91	38
	1	40	93	66
	1	42	93	94
	1	36	93	122
	1	38	92	149
	1	36	93	176
	1	37	95	202
	1	37	95	227
	2	45	100	27
	2	49	95	56
	2	50	86	82
	2	47	84	108
	2	46	86	134
	2	44	89	158
	2	44	90	182
	2	49	89	207
	2	49	89	230
	3	45	85	9
	3	42	88	38
	3	42	91	67
	3	42	90	93
	3	42	91	118
	3	43	91	144
	3	43	92	170
	3	40	93	195
	3	37	95	220
	3	38	95	245
	4	36	84	22
	4	39	88	49
	4	37	89	76
	4	42	87	103
	4	39	88	129
	4	39	89	155
	4	41	89	182
	4	41	89	208
	4	40	89	234
10	1	49	95	6
	1	48	92	20
	1	47	91	35
	1	47	91	49
	1	49	91	64
	1	47	91	78
	1	48	91	93
	1	48	91	108
	1	48	90	121
	1	50	90	136
	1	50	90	150
	1	49	88	165
	1	48	90	180
	1	48	91	194
	1	48	91	209
	1	47	90	224
	1	48	91	239
	2	52	85	11
	2	50	84	22
	2	51	85	34
	2	49	84	45
	2	47	87	57
	2	47	84	69
	2	48	84	80
	2	47	84	92
	2	49	84	104
	2	50	84	115
	2	49	83	127
	2	50	84	139
	2	51	84	150
	2	52	85	163
	2	51	84	175
	2	52	85	188
	2	49	85	200
	2	47	84	212
	2	46	85	224
	2	45	85	235
	2	51	84	247
	3	51	79	5
	3	49	79	16
	3	50	80	27
	3	50	80	39
	3	51	79	51

Person #	Scene #	y	x	t
	3	49	80	63
	3	50	80	74
	3	49	79	86
	3	52	80	96
	3	51	80	108
	3	51	81	120
	3	52	81	131
	3	51	81	143
	3	50	81	155
	3	50	80	167
	3	49	80	178
	3	49	80	191
	3	49	80	203
	3	49	78	214
	3	50	79	226
	3	51	81	238
	4	49	60	5
	4	48	62	17
	4	45	63	29
	4	44	63	41
	4	45	64	54
	4	43	64	66
	4	44	63	80
	4	43	63	92
	4	44	62	105
	4	44	63	118
	4	44	63	131
	4	43	63	144
	4	45	63	157
	4	44	63	170
	4	43	62	183
	4	43	63	196
	4	45	63	210
	4	43	63	223
	4	44	63	237
11	1	45	79	2
	1	45	80	24
	1	48	79	47
	1	52	80	69
	1	53	80	92
	1	52	79	115
	1	52	78	137
	1	52	78	159
	1	51	78	181
	1	52	79	202
	1	53	80	224
	1	54	80	245
	2	39	84	6
	2	40	86	29
	2	41	88	51
	2	38	89	72
	2	43	90	93
	2	47	87	114
	2	43	88	134
	2	42	91	154
	2	44	89	174
	2	43	89	215
	3	37	87	7
	3	40	89	28
	3	45	90	49
	3	43	93	72
	3	47	93	94
	3	49	92	116
	3	48	93	138
	3	50	94	160
	3	44	92	182
	3	46	91	203
	3	47	92	224
	3	48	94	245
	4	26	86	2
	4	15	89	25
	4	15	88	49
	4	25	88	70
	4	20	88	94
	4	20	88	117
	4	24	90	141
	4	24	92	164
	4	25	93	187
	4	27	91	210
	4	30	90	233

Person #	Scene #	y	x	t
12	1	52	89	1
	1	54	89	31
	1	53	90	62
	1	53	89	94
	1	53	90	126
	1	51	89	157
	1	49	88	190
	1	49	88	220
	2	45	93	24
	2	41	98	57
	2	46	98	88
	2	48	95	120
	2	44	100	150
	2	36	106	180
	2	45	96	210
	2	43	100	239
	3	49	95	25
	3	52	98	57
	3	53	98	90
	3	53	98	123
	3	55	97	153
	3	56	97	185
	3	54	98	216
	3	53	98	247
	4	40	92	2
	4	33	92	27
	4	29	93	62
	4	29	95	99
	4	30	94	134
	4	30	95	168
	4	31	92	203
	4	30	93	238
13	1	44	87	22
	1	42	86	57
	1	45	87	89
	1	46	88	124
	1	48	89	158
	1	43	90	190
	1	44	90	224
	2	33	53	20
	2	43	65	55
	2	44	64	90
	2	39	61	124
	2	38	61	156
	2	41	68	189
	2	37	64	222
	3	30	88	29
	3	30	88	64
	3	31	90	98
	3	29	90	132
	3	30	92	166
	3	31	92	201
	3	29	91	235
14	1	46	91	19
	1	43	87	56
	1	47	87	97
	1	48	88	134
	1	46	88	172
	1	47	89	211
	1	47	91	249
	2	42	83	13
	2	44	90	50
	2	49	87	88
	2	46	89	127
	2	43	92	165
	2	49	88	205
	2	47	90	244
	3	46	83	35
	3	47	84	73
	3	48	85	110
	3	45	84	149
	3	45	85	187
	3	48	84	225
	4	37	72	19
	4	35	73	51
	4	34	75	84
	4	34	76	117
	4	35	78	151
	4	35	78	184
	4	35	78	217

Figure A.11: Hanclapping Coordinates.

Person #	Scene #	y	x	t
15	1	48	84	17
	1	51	83	41
	1	53	84	66
	1	52	83	91
	1	51	83	116
	1	52	84	140
	1	52	85	164
	1	54	82	187
	1	54	84	210
	1	56	85	232
	2	43	88	12
	2	47	87	36
	2	42	88	59
	2	50	85	83
	2	49	86	107
	2	49	86	132
	2	47	87	155
	2	44	88	179
	2	49	85	204
	2	49	87	228
	3	50	85	12
	3	49	87	38
	3	52	87	62
	3	51	85	87
	3	49	86	135
	3	48	88	158
	3	47	88	182
	3	48	87	206
	3	48	88	231
	4	35	73	10
	4	35	73	34
	4	34	73	57
	4	37	72	81
	4	37	73	105
	4	35	73	127
	4	34	73	150
	4	36	73	173
	4	36	74	196
	4	37	73	219
	4	39	73	241
16	1	52	74	14
	1	50	73	47
	1	51	73	81
	1	51	74	114
	1	53	75	149
	1	53	76	183
	1	53	74	217
	2	43	89	2
	2	41	98	33
	2	45	93	66
	2	47	93	99
	2	45	99	132
	2	43	98	166
	2	46	95	201
	2	44	98	235
	3	49	86	10
	3	43	87	44
	3	47	88	76
	3	48	87	76
	3	45	88	109
	3	46	87	142
	3	46	87	142
	3	43	88	176
	3	49	89	208
	3	49	88	241
	4	43	99	10
	4	40	98	36
	4	38	100	63
	4	40	100	91
	4	38	101	120
	4	41	100	150
	4	38	99	178
	4	39	101	206
	4	35	104	234
17	1	31	86	22
	1	32	91	47
	1	31	92	75
	1	27	93	103
	1	27	94	130
	1	38	94	157

Person #	Scene #	y	x	t
	1	39	92	184
	1	39	94	211
	1	38	96	238
	2	33	92	2
	2	33	94	31
	2	34	95	59
	2	33	98	88
	2	38	93	116
	2	35	94	145
	2	31	97	172
	2	37	95	201
	2	39	92	230
	3	37	88	11
	3	36	89	40
	3	36	90	71
	3	36	86	101
	3	38	90	131
	3	36	90	161
	3	36	88	191
	3	36	90	221
	4	16	91	14
	4	12	92	36
	4	19	92	58
	4	19	94	81
	4	18	94	105
	4	21	95	130
	4	23	95	155
	4	23	94	179
	4	23	95	204
	4	23	93	228
18	1	41	92	22
	1	41	91	56
	1	38	91	90
	1	40	98	125
	1	39	91	161
	1	39	90	197
	1	44	94	232
	2	40	83	19
	2	41	84	54
	2	41	84	91
	2	44	83	128
	2	36	85	164
	2	33	85	198
	2	38	85	234
	3	43	90	15
	3	43	90	57
	3	43	90	101
	3	44	91	145
	3	46	91	190
	3	44	90	232
	4	38	97	15
	4	39	98	58
	4	36	96	98
	4	38	96	141
	4	32	97	185
	4	35	97	229
19	1	37	92	12
	1	39	93	42
	1	38	94	71
	1	38	93	99
	1	37	94	128
	1	38	94	158
	1	39	93	187
	1	38	94	216
	1	37	94	246
	2	45	91	13
	2	38	95	39
	2	42	94	67
	2	42	92	95
	2	47	90	124
	2	47	90	153
	2	51	88	182
	2	50	88	210
	2	48	90	241
	3	31	91	8
	3	34	91	39
	3	33	92	70
	3	32	92	100
	3	34	90	132
	3	31	91	164

Person #	Scene #	y	x	t
	3	32	91	196
	3	33	92	228
	4	32	89	15
	4	33	89	38
	4	34	89	64
	4	32	87	89
	4	32	87	115
	4	32	86	140
	4	32	88	165
	4	31	87	190
	4	32	87	217
	4	33	88	243
2	1	38	78	16
	1	39	79	41
	1	42	79	64
	1	42	83	89
	1	40	82	113
	1	39	84	137
	1	43	84	161
	1	46	85	184
	1	45	84	206
	1	46	85	230
	2	41	86	12
	2	42	84	34
	2	46	84	59
	2	48	82	81
	2	45	82	103
	2	45	82	126
	2	40	84	148
	2	36	86	171
	2	43	85	193
	2	44	84	216
	2	46	84	239
	3	48	86	8
	3	50	83	33
	3	47	84	58
	3	46	82	81
	3	46	81	104
	3	44	83	129
	3	47	80	155
	3	48	83	181
	3	50	84	206
	3	49	83	230
	4	43	79	14
	4	42	79	39
	4	43	80	64
	4	42	80	89
	4	41	79	114
	4	38	79	139
	4	39	78	167
	4	39	79	194
	4	37	79	221
	4	44	78	247
20	1	49	82	1
	1	49	82	27
	1	48	81	55
	1	48	83	83
	1	48	83	114
	1	46	82	141
	1	47	83	169
	1	46	83	198
	1	45	82	227
	2	54	62	30
	2	50	60	59
	2	53	67	89
	2	52	66	117
	2	49	64	147
	2	47	60	176
	2	46	59	204
	2	50	61	233
	3	54	67	13
	3	56	66	40
	3	54	66	73
	3	54	68	106
	3	53	67	141
	3	50	68	174
	3	53	70	206
	3	55	68	240
	4	46	66	14
	4	47	66	46

Figure A.12: Handclapping Coordinates Continued.

Person #	Scene #	y	x	t
	4	49	64	79
	4	51	66	113
	4	49	64	144
	4	49	66	177
	4	49	66	210
	4	48	66	243
21	1	35	98	12
	1	34	96	41
	1	32	99	71
	1	29	97	105
	1	30	99	139
	1	28	97	173
	1	29	99	209
	1	27	98	242
	2	35	88	17
	2	35	89	55
	2	30	88	93
	2	33	87	130
	2	35	85	170
	2	27	86	208
	2	32	84	249
	3	36	88	11
	3	36	88	41
	3	35	86	71
	3	36	87	103
	3	34	86	132
	3	35	85	163
	3	35	86	195
	3	34	86	226
	4	22	73	15
	4	23	74	53
	4	22	74	91
	4	20	74	166
	4	21	74	202
	4	22	72	239
22	1	46	78	12
	1	47	78	37
	1	45	78	64
	1	47	78	89
	1	48	79	112
	1	50	80	137
	1	48	80	162
	1	49	81	186
	1	52	81	212
	1	50	81	236
	2	54	94	13
	2	51	92	40
	2	51	93	68
	2	46	94	98
	2	49	93	125
	2	47	93	153
	2	48	92	181
	2	51	91	209
	2	49	92	237
	3	51	98	10
	3	45	1⊞	41
	3	49	1⊞	71
	3	48	1⊞	101
	3	47	99	131
	3	47	99	160
	3	46	99	191
	3	41	99	221
	4	42	74	13
	4	38	75	45
	4	40	77	77
	4	42	77	108
	4	42	77	140
	4	40	77	173
	4	41	77	204
	4	38	77	237
23	1	45	79	15
	1	46	83	48
	1	42	86	79
	1	44	87	110
	1	44	88	141
	1	45	89	174
	1	45	89	206
	1	45	90	236
	2	53	88	17
	2	49	88	49

Person #	Scene #	y	x	t
	2	50	88	83
	2	47	88	116
	2	47	87	149
	2	49	87	182
	2	46	89	215
	2	45	88	249
	3	50	84	17
	3	48	86	53
	3	46	85	91
	3	43	84	126
	3	44	85	164
	3	44	85	201
	3	44	85	239
	4	34	53	19
	4	35	54	57
	4	33	51	97
	4	34	52	134
	4	34	53	173
	4	32	53	209
	4	31	55	247
24	1	46	85	10
	1	45	85	34
	1	44	87	58
	1	42	89	82
	1	42	90	106
	1	44	90	130
	1	43	90	156
	1	44	91	179
	1	45	91	203
	1	45	92	227
	2	52	93	11
	2	48	97	34
	2	47	94	57
	2	45	93	79
	2	46	95	103
	2	45	95	127
	2	44	97	150
	2	42	98	173
	2	47	95	196
	2	50	94	219
	2	50	93	242
	3	48	86	10
	3	49	87	32
	3	47	86	54
	3	48	86	77
	3	47	86	100
	3	46	86	122
	3	44	87	143
	3	46	88	166
	3	46	86	189
	3	46	86	210
	3	45	88	232
	4	45	82	9
	4	46	82	32
	4	44	82	53
	4	45	82	75
	4	45	81	98
	4	42	81	119
	4	43	82	140
	4	44	82	162
	4	44	83	185
	4	45	82	206
	4	45	82	228
25	1	43	80	12
	1	39	78	37
	1	39	79	62
	1	44	80	88
	1	45	81	114
	1	43	81	140
	1	47	82	165
	1	45	82	192
	1	43	82	217
	1	45	83	244
	2	41	87	13
	2	44	87	44
	2	44	86	73
	2	46	86	103
	2	43	88	134
	2	38	90	164
	2	40	88	195

Person #	Scene #	y	x	t
	2	43	87	226
	3	50	84	10
	3	46	85	86
	3	51	85	62
	3	47	86	88
	3	46	85	115
	3	47	85	143
	3	47	85	169
	3	47	85	197
	3	48	86	225
	4	49	84	15
	4	44	85	46
	4	44	85	78
	4	46	84	109
	4	50	84	141
	4	48	84	173
	4	44	84	203
	4	45	83	234
3	1	44	78	11
	1	42	78	41
	1	44	81	70
	1	42	84	100
	1	41	81	129
	1	41	82	159
	1	39	85	189
	1	36	93	219
	2	54	77	11
	2	54	78	39
	2	49	76	67
	2	53	76	94
	2	49	76	122
	2	47	77	151
	2	48	79	179
	2	49	80	208
	2	47	76	237
	3	54	82	10
	3	53	81	38
	3	52	79	67
	3	49	79	96
	3	48	81	127
	3	49	80	183
	3	50	79	211
	3	45	77	241
	4	43	67	10
	4	49	66	37
	4	43	64	66
	4	48	65	94
	4	48	66	124
	4	44	66	153
	4	48	74	182
	4	48	72	211
	4	45	72	241
4	1	50	93	11
	1	48	91	38
	1	52	91	65
	1	51	93	91
	1	52	91	118
	1	52	92	144
	1	52	92	169
	1	53	91	192
	1	52	92	221
	1	53	92	247
	2	55	86	11
	2	53	84	41
	2	54	84	69
	2	52	84	97
	2	52	84	125
	2	53	85	152
	2	53	84	178
	2	54	84	205
	2	55	83	231
	3	53	91	7
	3	56	91	32
	3	55	92	58
	3	55	91	84
	3	56	89	110
	3	55	90	135
	3	54	92	162
	3	58	91	188
	3	58	91	214

Figure A.13: Handclapping Coordinates Continued.

Person #	Scene #	y	x	t
	3	56	89	239
	4	42	64	10
	4	45	64	39
	4	47	63	68
	4	46	63	96
	4	47	63	124
	4	47	64	152
	4	45	64	180
	4	40	62	208
	4	46	62	237
5	1	47	88	10
	1	46	90	29
	1	46	88	46
	1	47	90	65
	1	46	90	84
	1	46	90	102
	1	45	89	119
	1	44	90	137
	1	45	90	156
	1	45	92	174
	1	45	91	192
	1	46	91	210
	1	46	91	228
	1	45	90	246
	2	48	87	11
	2	49	91	29
	2	46	94	45
	2	44	95	62
	2	45	91	77
	2	47	91	94
	2	46	89	110
	2	47	88	127
	2	48	89	144
	2	46	92	161
	2	41	94	176
	2	42	94	192
	2	49	88	209
	2	50	89	224
	2	46	90	241
	3	48	84	7
	3	49	84	20
	3	48	80	34
	3	48	83	51
	3	49	84	67
	3	48	84	83
	3	48	85	100
	3	48	84	115
	3	47	87	132
	3	48	85	148
	3	47	85	164
	3	48	85	181
	3	49	85	198
	3	48	86	215
	3	48	85	231
	3	47	85	248
	4	46	75	7
	4	48	76	24
	4	48	76	40
	4	48	76	59
	4	46	78	76
	4	46	77	92
	4	45	80	110
	4	45	78	126
	4	44	79	142
	4	45	79	158
	4	44	80	175
	4	44	80	191
	4	45	80	208
	4	44	80	224
	4	43	80	240
6	1	45	88	15
	1	45	88	43
	1	44	91	71
	1	45	92	98
	1	45	93	126
	1	46	92	154
	1	45	93	182
	1	46	92	210
	1	46	92	239
	2	48	113	9

Person #	Scene #	y	x	t
	2	43	106	34
	2	47	104	56
	2	46	105	81
	2	43	111	104
	2	39	109	127
	2	38	108	150
	2	44	104	174
	2	44	104	197
	2	44	102	219
	2	46	98	244
	3	52	97	9
	3	44	97	36
	3	49	96	60
	3	50	98	84
	3	49	97	108
	3	52	97	131
	3	51	96	155
	3	51	97	178
	3	50	96	202
	3	52	97	224
	3	52	98	248
	4	55	80	10
	4	55	78	32
	4	56	80	55
	4	54	82	75
	4	51	81	98
	4	52	83	120
	4	51	81	142
	4	53	82	163
	4	55	83	184
	4	53	82	207
	4	53	82	228
7	1	40	85	18
	1	35	88	49
	1	40	89	81
	1	44	91	113
	1	41	90	144
	1	33	90	175
	1	37	89	205
	1	35	90	235
	2	29	83	19
	2	27	84	53
	2	33	84	88
	2	18	88	124
	2	29	88	158
	2	37	87	194
	2	33	87	232
	3	31	105	17
	3	32	105	49
	3	36	103	81
	3	35	104	115
	3	33	105	149
	3	34	108	215
	3	32	109	248
	4	41	90	21
	4	40	89	63
	4	44	87	102
	4	46	86	139
	4	46	87	176
	4	46	86	213
	4	46	87	249
8	1	27	84	12
	1	30	84	41
	1	32	85	70
	1	34	85	96
	1	33	84	124
	1	30	84	150
	1	30	85	177
	1	34	85	204
	1	33	85	231
	2	36	77	13
	2	41	79	39
	2	38	79	65
	2	37	82	91
	2	32	82	117
	2	38	79	142
	2	43	77	168
	2	38	78	194
	2	40	79	220
	2	34	77	246

Person #	Scene #	y	x	t
	3	38	88	8
	3	41	89	32
	3	40	86	58
	3	40	85	85
	3	42	85	111
	3	41	84	137
	3	39	84	162
	3	39	85	195
	3	39	84	220
	3	36	85	247
	4	34	72	34
	4	37	71	61
	4	34	74	88
	4	34	73	114
	4	27	72	166
	4	34	74	192
	4	35	72	218
	4	34	73	245
9	1	33	86	9
	1	30	84	33
	1	30	85	56
	1	32	83	80
	1	33	83	104
	1	33	83	128
	1	35	84	152
	1	36	82	176
	1	37	83	202
	1	37	82	225
	2	30	83	9
	2	36	82	34
	2	42	82	55
	2	42	81	78
	2	42	81	102
	2	42	81	126
	2	37	82	150
	2	33	80	174
	2	35	80	198
	2	37	81	222
	2	41	81	247
	3	37	84	11
	3	36	84	30
	3	36	82	50
	3	35	82	72
	3	35	84	94
	3	38	83	117
	3	38	83	140
	3	36	84	162
	3	36	81	184
	3	33	82	207
	3	33	82	230
	4	36	63	10
	4	35	62	33
	4	34	62	54
	4	36	62	77
	4	34	61	99
	4	33	62	123
	4	31	62	145
	4	33	61	168
	4	33	61	192
	4	33	61	214
	4	32	60	238

Figure A.14: Handclapping Coordinates Continued.

Person #	Scene #	y	x	t
1	1	15	95	13
	1	13	95	52
	1	12	94	89
	1	11	95	126
	1	11	97	161
	1	14	95	197
	1	15	98	231
	2	16	94	30
	2	15	95	63
	2	21	95	96
	2	28	91	130
	2	25	92	163
	2	10	99	195
	2	23	93	228
	3	25	85	19
	3	25	96	58
	3	23	86	93
	3	25	88	129
	3	21	89	165
	3	24	88	199
	3	22	90	233
	4	14	82	26
	4	13	83	68
	4	14	85	109
	4	17	84	149
	4	15	84	188
	4	17	83	224
10	1	13	86	21
	1	14	88	61
	1	15	87	103
	1	14	86	141
	1	18	88	179
	1	17	87	220
	2	3	79	16
	2	11	77	48
	2	22	78	81
	2	22	80	118
	2	17	80	155
	2	9	80	196
	2	4	78	235
	3	15	72	19
	3	12	71	57
	3	12	70	92
	3	12	69	128
	3	13	69	168
	3	11	70	208
	3	8	70	247
	4	10	66	15
	4	8	66	50
	4	9	65	86
	4	10	65	122
	4	9	65	159
	4	8	66	198
	4	9	64	237
11	1	24	80	20
	1	24	76	52
	1	26	79	86
	1	28	80	124
	1	28	80	165
	1	27	82	208
	1	27	82	247
	2	10	103	21
	2	30	98	55
	2	21	102	88
	2	14	105	122
	2	30	97	187
	2	27	98	220
	3	21	92	18
	3	14	86	52
	3	15	92	84
	3	15	90	116
	3	19	89	149
	3	14	92	180
	3	14	92	212
	3	17	94	244
	4	3	77	2
	4	3	80	37
	4	5	81	72
	4	6	80	107
	4	6	82	140

Person #	Scene #	y	x	t
	4	6	81	173
	4	2	83	207
12	1	20	81	23
	1	17	85	83
	1	20	88	143
	1	23	89	197
	2	16	101	25
	2	33	96	79
	2	19	105	134
	2	28	105	184
	2	18	109	243
	3	22	88	23
	3	18	90	84
	3	18	91	144
	3	16	91	207
	4	8	81	15
	4	6	86	72
	4	5	85	128
	4	6	89	187
	4	6	93	246
13	1	11	96	20
	1	8	96	75
	1	10	100	124
	1	13	102	170
	1	12	104	213
	2	22	98	2
	2	26	99	39
	2	17	106	78
	2	29	98	116
	2	15	107	153
	2	29	102	190
	2	19	108	227
	3	22	88	19
	3	22	88	60
	3	19	90	104
	3	19	92	146
	3	21	92	186
	3	21	92	224
	4	9	84	33
	4	5	85	94
	4	6	85	152
	4	6	87	206
14	1	18	86	2
	1	19	89	48
	1	17	89	95
	1	15	93	141
	1	14	92	187
	1	13	93	232
	2	33	94	23
	2	26	97	73
	2	27	94	121
	2	29	97	169
	2	17	101	217
	3	24	95	24
	3	25	94	72
	3	25	95	119
	3	25	95	166
	3	25	97	212
	4	12	63	19
	4	10	64	66
	4	9	66	111
	4	10	66	154
	4	11	67	199
	4	9	66	243
15	1	26	80	18
	1	26	81	68
	1	26	81	115
	1	29	81	167
	1	30	82	215
	2	26	88	44
	2	26	90	95
	2	8	92	147
	2	13	93	201
	3	17	85	15
	3	14	87	58
	3	17	87	105
	3	16	87	152
	3	20	85	197
	3	21	85	244
	4	6	58	15

Person #	Scene #	y	x	t
	4	7	60	68
	4	9	62	120
	4	9	62	174
	4	10	66	228
16	1	21	91	16
	1	20	91	52
	1	21	92	89
	1	24	91	125
	1	26	93	158
	1	27	95	193
	1	33	95	229
	2	15	96	20
	2	18	97	55
	2	19	97	91
	2	20	99	125
	2	11	106	160
	2	22	102	196
	2	32	99	228
	3	20	80	18
	3	22	81	57
	3	19	81	94
	3	21	81	132
	3	21	81	169
	3	24	84	205
	3	22	85	242
	4	2	83	120
17	1	15	90	15
	1	9	93	43
	1	5	93	72
	1	8	94	101
	1	8	94	130
	1	6	96	157
	1	8	95	186
	1	8	93	214
	1	8	94	244
	2	2	107	18
	2	11	108	53
	2	5	110	87
	2	13	104	121
	2	15	105	153
	2	11	109	185
	2	10	110	219
	3	10	80	19
	3	9	81	54
	3	11	81	87
	3	8	79	121
	3	8	81	155
	3	9	83	187
	3	6	82	219
	4	12	80	12
	4	4	85	39
	4	8	88	68
	4	6	89	95
	4	5	88	123
	4	4	88	149
	4	4	86	176
	4	4	88	205
	4	3	88	231
18	1	3	90	23
	1	4	91	71
	1	3	94	115
	1	3	95	162
	1	5	93	208
	2	1	100	42
	2	5	97	90
	2	1	99	136
	2	14	96	184
	2	3	98	237
	3	4	80	29
	3	5	81	80
	3	4	84	130
	3	5	85	180
	3	6	84	228
	4	9	96	51
	4	8	99	101
	4	9	99	152
	4	10	98	204
19	1	21	94	20
	1	21	95	66
	1	19	94	117

Figure A.15: Handwaving Coordinates.

Person #	Scene #	y	x	t
	1	18	93	166
	1	20	94	217
	2	1	100	42
	2	5	97	90
	2	1	99	136
	2	14	96	184
	2	3	98	237
	3	13	83	16
	3	10	85	61
	3	11	89	105
	3	11	88	151
	3	9	89	198
	3	11	92	244
	4	11	82	29
	4	9	83	83
	4	12	85	132
	4	14	83	176
	4	11	85	221
2	1	9	75	21
	1	12	77	60
	1	6	73	97
	1	11	75	137
	1	7	77	241
	2	32	76	42
	2	12	74	71
	2	12	73	99
	2	17	74	126
	2	19	72	154
	2	18	76	178
	2	17	71	202
	2	19	74	225
	3	19	75	15
	3	26	76	44
	3	24	68	73
	3	29	72	105
	3	32	76	136
	3	29	72	136
	3	29	72	167
	3	35	69	199
	3	32	77	227
	4	11	72	20
	4	9	62	54
	4	11	63	88
	4	16	59	124
	4	18	66	158
	4	19	69	190
	4	15	70	220
	4	18	72	249
20	1	4	83	16
	1	4	82	59
	1	5	82	103
	1	7	83	149
	1	7	83	195
	1	7	81	241
	2	1	71	17
	2	9	71	53
	2	17	73	89
	2	11	71	125
	2	3	73	162
	2	5	71	198
	2	16	73	235
	3	8	70	18
	3	6	70	56
	3	6	70	95
	3	9	70	133
	3	9	70	172
	3	7	70	212
	4	4	73	18
	4	2	72	55
	4	2	69	131
	4	2	69	170
	4	3	70	209
	4	2	70	248
21	1	11	100	20
	1	4	102	66
	1	8	91	111
	1	9	99	154
	1	6	99	209
	2	6	90	23
	2	7	89	182

Person #	Scene #	y	x	t
	2	2	89	230
	3	5	91	18
	3	3	89	61
	3	5	96	100
	3	8	94	137
	3	9	94	176
	3	6	92	213
	4	4	66	21
	4	6	70	164
	4	13	59	211
22	1	21	73	17
	1	19	74	64
	1	18	76	108
	1	17	73	150
	1	12	73	193
	1	12	76	236
	2	12	85	22
	2	18	85	106
	2	15	85	149
	2	7	88	196
	2	5	85	241
	3	7	90	20
	3	10	96	65
	3	7	95	110
	3	6	98	154
	3	5	99	199
	3	5	100	245
	4	11	75	25
	4	10	77	78
	4	10	76	132
	4	12	77	178
	4	12	78	224
23	1	5	84	18
	1	5	87	60
	1	3	86	103
	1	5	87	145
	1	6	88	187
	1	8	87	231
	2	5	89	61
	2	16	88	101
	2	9	89	142
	2	4	89	183
	2	6	89	225
	3	8	86	20
	3	7	89	62
	3	8	88	104
	3	9	90	143
	3	11	89	184
	3	9	90	224
	4	3	58	25
	4	6	54	70
	4	5	56	113
	4	8	55	158
	4	7	53	203
	4	5	54	244
24	1	10	89	20
	1	12	93	67
	1	14	97	113
	1	11	94	158
	1	11	92	201
	1	12	91	244
	2	25	87	22
	2	20	89	69
	2	11	93	113
	2	21	88	160
	2	29	89	206
	3	8	86	19
	3	9	89	63
	3	10	88	105
	3	11	87	145
	3	12	90	190
	3	11	88	233
	4	5	80	19
	4	9	84	66
	4	6	80	109
	4	8	81	153
	4	9	83	198
	4	8	84	241
25	1	8	77	18
	1	10	77	57

Person #	Scene #	y	x	t
	1	10	80	98
	1	11	81	137
	1	11	80	178
	1	12	80	217
	2	5	81	16
	2	6	83	53
	2	19	83	92
	2	10	85	130
	2	15	84	167
	2	21	85	205
	2	16	84	242
	3	11	83	19
	3	12	81	62
	3	11	82	105
	3	13	81	149
	3	12	82	192
	3	12	83	235
	4	6	80	19
	4	9	81	59
	4	7	81	100
	4	8	82	140
	4	8	79	180
	4	9	79	218
3	1	9	78	19
	1	11	80	65
	1	13	84	112
	1	13	81	158
	1	13	78	201
	1	12	79	246
	2	13	78	23
	2	5	74	65
	2	18	79	107
	2	4	81	151
	2	14	78	193
	2	25	81	235
	3	25	75	22
	3	21	68	68
	3	26	64	113
	3	22	73	158
	3	17	75	203
	4	9	75	20
	4	8	74	64
	4	8	72	108
	4	11	75	150
	4	11	85	193
	4	6	78	236
4	1	9	87	17
	1	9	88	55
	1	6	89	92
	1	9	89	130
	1	9	88	166
	1	12	91	202
	1	11	91	238
	2	6	82	19
	2	16	81	61
	2	11	83	101
	2	12	81	143
	2	8	82	187
	2	8	82	230
	3	18	90	18
	3	14	92	56
	3	15	91	92
	3	14	93	129
	3	12	95	166
	3	14	92	205
	3	14	91	243
	4	3	64	22
	4	3	64	67
	4	4	65	151
	4	4	65	193
	4	6	66	234
5	1	19	83	15
	1	23	85	52
	1	18	85	85
	1	20	84	118
	1	16	84	153
	1	14	87	190
	1	18	86	227
	2	34	89	13
	2	31	84	49

Figure A.16: Handwaving Coordinates Continued.

Person #	Scene #	y	x	t
	2	26	86	81
	2	40	88	116
	2	40	86	149
	2	38	86	180
	2	34	87	211
	2	20	87	244
	3	34	83	13
	3	33	78	77
	3	32	78	77
	3	32	78	108
	3	30	80	141
	3	30	76	172
	3	40	85	205
	3	35	80	237
	4	19	76	14
	4	19	75	48
	4	14	75	118
	4	16	74	151
	4	18	77	183
	4	14	75	215
	4	17	76	247
6	1	16	102	12
	1	9	100	51
	1	11	100	90
	1	13	104	129
	1	13	103	168
	1	14	104	207
	1	16	108	248
	2	6	112	17
	2	8	112	51
	2	16	108	86
	2	20	110	120
	2	18	110	154
	2	11	108	187
	2	13	109	224
	3	9	89	17
	3	13	89	55
	3	14	88	90
	3	14	90	125
	3	14	89	162
	3	14	91	196
	3	13	90	232
	4	7	81	17
	4	7	84	54
	4	7	83	90
	4	10	82	127
	4	9	83	165
	4	10	86	202
	4	10	84	237
7	1	11	90	39
	1	12	93	119
	1	11	93	194
	2	9	100	38
	2	19	96	105
	2	5	103	180
	2	22	98	247
	3	16	99	18
	3	19	96	71
	3	17	97	125
	3	16	100	181
	3	17	98	236
	4	10	84	22
	4	9	83	103
	4	15	81	185
8	1	20	78	19
	1	17	82	57
	1	17	80	93
	1	16	78	128
	1	19	81	166
	1	18	81	203
	1	16	78	242
	2	20	87	18
	2	34	85	54
	2	28	83	90
	2	22	81	123
	2	38	81	156
	2	36	79	189
	2	32	81	222
	3	19	86	16
	3	24	84	47

Person #	Scene #	y	x	t
	3	20	86	78
	3	18	86	108
	3	17	86	141
	3	25	86	173
	3	22	85	205
	3	17	83	236
	4	17	69	15
	4	13	69	49
	4	17	65	82
	4	16	65	114
	4	16	65	150
	4	18	62	186
	4	15	63	222
9	1	19	81	20
	1	14	81	57
	1	16	80	94
	1	16	78	129
	1	18	80	166
	1	17	80	203
	1	15	80	242
	2	22	87	19
	2	34	83	55
	2	28	83	90
	2	21	81	124
	2	37	80	157
	2	36	80	190
	2	31	82	222
	3	16	19	85
	3	47	24	83
	3	78	19	86
	3	108	18	86
	3	142	18	86
	3	173	23	85
	3	205	20	84
	3	237	20	82
	4	15	17	68
	4	49	11	67
	4	82	17	65
	4	115	16	64
	4	150	15	64
	4	186	17	61
	4	222	14	62

Figure A.17: Handwaving Coordinates Continued.

Person #	Scene #	y	x	t
1	1	20	78	108
	1	30	70	65
	1	143	78	71
	1	154	77	120
	2	16	76	131
	2	27	76	92
	2	37	60	61
	2	48	63	31
	2	130	47	24
	2	141	62	50
	2	151	61	80
	2	161	78	112
	3	21	73	86
	3	32	75	31
	3	139	75	72
	3	150	78	128
	4	28	75	112
	4	38	76	76
	4	49	75	33
	4	133	74	59
	4	144	76	103
	4	222	73	119
	4	233	77	73
	4	243	78	29
10	1	71	120	15
	1	69	82	24
	1	76	38	34
	1	84	65	157
	1	87	102	165
	1	88	142	175
	1	71	136	247
	2	41	71	74
	2	58	52	84
	2	55	32	91
	2	54	35	195
	2	61	61	204
	2	38	78	212
	3	73	128	21
	3	69	86	29
	3	74	30	39
	3	83	39	150
	3	78	79	157
	3	88	137	167
	3	76	100	233
	3	75	54	241
	4	73	53	25
	4	84	93	35
	4	77	127	43
	4	74	130	124
	4	80	85	134
	4	81	50	142
	4	81	19	151
	4	76	43	198
	4	80	77	206
	4	75	111	214
	4	86	143	223
11	1	74	122	16
	1	81	88	26
	1	82	51	36
	1	72	28	132
	1	78	68	143
	1	82	102	152
	1	77	141	162
	1	75	120	248
	2	61	150	14
	2	69	118	26
	2	76	101	34
	2	86	65	46
	2	92	35	55
	2	96	37	141
	2	84	104	163
	2	69	129	172
	3	91	125	15
	3	86	96	24
	3	84	60	34
	3	81	22	44
	3	81	42	146
	3	84	89	156
	3	89	137	166
	4	69	120	21
	4	77	75	31
	4	74	33	122
	4	78	87	133
	4	86	139	143
	4	81	133	233
	4	74	93	242

Person #	Scene #	y	x	t
12	1	60	131	45
	1	63	68	57
	1	63	23	67
	1	75	51	144
	1	75	100	154
	2	43	131	28
	2	58	109	39
	2	65	78	50
	2	74	48	60
	2	80	24	154
	2	75	57	163
	2	66	92	175
	2	61	113	184
	2	53	135	194
	2	52	155	205
	3	78	133	17
	3	76	86	27
	3	76	41	37
	3	90	41	148
	3	93	83	157
	3	90	134	168
	3	76	113	249
	4	76	99	41
	4	78	50	52
	4	77	79	142
	4	75	126	153
	4	85	96	220
	4	83	50	231
	4	83	12	242
13	1	71	73	27
	1	68	34	37
	1	72	22	135
	1	75	110	155
	2	44	142	14
	2	55	112	25
	2	57	79	25
	2	66	45	45
	2	63	42	196
	2	53	73	206
	2	50	105	216
	2	43	128	225
	3	69	116	20
	3	71	68	31
	3	73	28	40
	3	74	36	152
	3	74	84	162
	3	80	123	171
	4	71	119	33
	4	71	74	43
	4	73	98	161
	4	75	154	173
14	1	70	133	17
	1	65	46	38
	1	72	80	195
	2	50	138	31
	2	53	105	43
	2	68	76	53
	2	72	29	65
	2	82	57	189
	2	61	98	201
	2	66	129	213
	3	78	131	43
	3	73	79	53
	3	71	25	65
	3	66	82	194
	3	71	143	206
	4	77	105	29
	4	76	56	40
	4	79	35	119
	4	70	80	129
	4	74	131	141
15	1	73	114	33
	1	74	76	44
	1	70	42	54
	1	73	44	162
	1	74	79	172
	1	75	119	183
	2	86	46	20
	2	82	78	30
	2	71	110	41
	2	71	131	51
	2	64	144	184
	2	67	121	194
	2	78	99	204
	2	82	70	214
	2	94	38	225

Person #	Scene #	y	x	t
	3	76	88	24
	3	75	39	34
	3	74	54	147
	3	73	102	157
	4	83	96	38
	4	83	57	49
	4	78	21	61
	4	87	27	135
	4	83	75	147
	4	80	111	157
	4	81	101	242
16	1	69	129	23
	1	68	93	32
	1	69	48	42
	1	67	34	139
	1	71	75	149
	1	69	116	158
	2	63	147	32
	2	62	123	42
	2	72	95	52
	2	72	63	62
	2	82	33	71
	2	68	77	168
	2	70	104	178
	2	59	130	188
	3	73	133	25
	3	70	83	34
	3	67	35	44
	3	72	74	173
	3	73	120	182
	4	79	122	26
	4	74	84	35
	4	71	42	45
	4	75	27	131
	4	74	66	141
	4	73	105	151
17	1	68	119	18
	1	69	74	29
	1	71	24	40
	1	73	78	190
	2	56	135	17
	2	45	111	26
	2	35	60	49
	2	32	29	158
	2	41	71	181
	2	49	117	200
	3	77	123	16
	3	73	75	26
	3	72	27	37
	3	73	67	153
	3	75	115	163
	4	70	132	24
	4	74	84	35
	4	75	31	46
	4	72	80	122
	4	74	130	132
	4	72	125	194
	4	74	26	215
18	1	70	29	18
	1	68	60	27
	1	69	89	36
	1	70	125	45
	1	71	128	102
	1	68	96	112
	1	68	59	122
	1	65	31	131
	1	67	28	239
	2	47	31	21
	2	62	106	141
	2	47	53	160
	3	74	59	27
	4	78	59	31
	4	83	99	42
	4	78	110	132
	4	78	67	142
	4	77	36	151
	4	82	48	198
	4	79	76	206
	4	79	109	215
19	1	83	140	17
	1	78	109	25
	1	77	79	34
	1	76	40	44
	1	72	23	120
	1	74	63	130
	1	72	95	139

Figure A.18: Jogging Coordinates.

Person #	Scene #	y	x	t
	1	77	131	149
	1	75	112	209
	1	72	77	218
	1	71	40	228
	2	41	27	23
	2	58	67	42
	2	79	117	62
	2	78	106	163
	2	56	59	182
	2	40	27	201
	3	78	126	14
	3	74	89	22
	3	73	49	32
	3	76	51	100
	3	77	98	110
	3	76	142	120
	3	86	133	190
	3	82	92	199
	3	76	57	207
	4	80	118	25
	4	74	88	33
	4	75	51	43
	4	78	31	121
	4	83	77	132
	4	78	113	141
	4	75	132	208
	4	73	45	227
2	1	20	75	112
	1	30	73	66
	1	131	75	52
	1	141	76	96
	1	212	70	91
	1	221	76	51
	2	117	59	99
	2	127	56	117
	2	138	74	138
	2	197	68	114
	2	207	77	92
	2	219	50	72
	2	229	59	57
	2	239	36	41
	3	70	41	20
	3	74	77	31
	3	75	112	42
	3	78	144	53
	3	76	125	100
	3	75	92	109
	3	75	55	119
	3	79	22	129
	3	73	38	221
	3	75	69	230
	3	76	112	242
	4	29	103	73
	4	40	73	68
	4	51	73	36
	4	102	69	56
	4	113	71	90
	4	123	76	121
	4	202	77	112
	4	212	72	83
	4	223	80	53
20	1	71	41	16
	1	73	78	25
	1	73	119	34
	1	75	119	201
	1	74	83	210
	1	75	40	220
	2	14	10	20
	2	31	23	30
	2	27	41	39
	2	43	60	48
	2	44	87	57
	2	66	121	67
	2	69	146	137
	2	62	108	146
	2	41	79	156
	2	24	40	174
	2	28	25	184
	2	15	13	193
	3	76	58	17
	3	74	105	25
	3	77	128	111
	3	73	78	120
	3	70	32	128
	3	67	35	213
	3	74	86	223

Person #	Scene #	y	x	t
	3	77	139	233
	4	79	57	29
	4	78	96	37
	4	75	132	46
	4	81	111	115
	4	75	76	122
	4	78	30	131
	4	78	59	191
	4	79	105	200
	4	76	142	209
21	1	70	154	7
	1	70	111	17
	1	72	54	29
	1	71	3	40
	1	60	36	191
	1	62	89	202
	1	65	142	212
	2	22	154	20
	2	17	140	30
	2	32	126	42
	2	33	105	53
	2	47	85	63
	2	51	53	74
	2	73	19	86
	2	69	15	150
	2	65	52	161
	2	42	83	172
	2	45	104	183
	2	24	125	194
	2	30	139	206
	2	16	155	216
	3	72	153	8
	3	74	91	19
	3	72	33	30
	3	74	56	170
	3	78	119	181
	4	84	40	40
	4	79	74	49
	4	76	123	61
	4	77	144	173
	4	83	90	186
	4	85	38	197
22	1	54	113	16
	1	52	61	27
	1	59	51	161
	1	66	111	173
	2	56	128	17
	2	42	97	28
	2	29	39	49
	2	42	79	192
	2	49	114	204
	3	64	111	18
	3	63	56	30
	3	67	54	181
	3	67	103	192
	4	74	71	35
	4	70	116	45
	4	80	131	119
	4	73	89	129
	4	79	41	140
	4	77	38	211
	4	77	83	221
	4	70	128	231
23	1	81	99	19
	1	76	50	29
	1	72	32	120
	1	75	78	130
	1	78	132	140
	1	71	123	241
	2	33	36	46
	2	52	68	65
	2	82	120	85
	2	71	109	148
	2	46	64	167
	3	68	116	15
	3	71	48	27
	3	64	63	94
	3	70	124	104
	3	73	122	204
	4	80	111	26
	4	78	76	35
	4	72	43	44
	4	76	67	96
	4	78	102	106
	4	74	132	115
	4	81	132	185

Person #	Scene #	y	x	t
	4	76	99	194
	4	78	66	203
24	1	76	121	17
	1	70	28	181
	1	75	67	191
	1	79	111	202
	2	40	136	37
	2	52	109	56
	2	76	63	78
	2	69	70	202
	2	52	109	222
	2	38	137	242
	3	74	57	31
	3	72	55	149
	3	70	135	169
	4	78	60	34
	4	77	87	132
	4	79	50	143
	4	75	24	152
	4	78	52	216
	4	76	85	226
	4	76	119	237
25	1	80	122	23
	1	78	77	33
	1	74	31	43
	1	80	77	109
	1	84	119	118
	1	76	96	173
	1	73	57	182
	2	47	18	25
	2	56	50	44
	2	70	95	63
	2	81	111	146
	2	65	58	167
	2	54	22	186
	3	72	32	106
	3	70	72	115
	3	77	86	202
	4	72	49	32
	4	71	84	41
	4	78	128	52
	4	75	108	149
	4	80	60	160
	4	79	65	225
	4	75	104	234
3	1	68	141	12
	1	66	105	21
	1	67	62	32
	1	69	23	42
	1	65	35	140
	1	66	70	149
	1	69	118	160
	2	32	153	22
	2	27	146	30
	2	34	132	42
	2	32	120	52
	2	43	106	63
	2	42	93	73
	2	54	80	83
	2	54	58	94
	2	62	40	103
	2	60	17	184
	2	59	39	193
	2	47	61	204
	2	50	85	216
	2	36	101	226
	2	35	117	237
	3	70	96	21
	3	67	48	31
	3	69	47	116
	3	68	90	125
	4	80	37	23
	4	82	78	33
	4	79	120	44
	4	79	121	133
	4	80	75	144
	4	84	36	155
	4	79	66	219
	4	76	106	229
4	1	72	103	21
	1	75	54	32
	1	79	71	110
	1	78	114	120
	1	71	112	218
	1	70	65	229
	2	47	35	66

Figure A.19: Jogging Coordinates Continued.

Person #	Scene #	y	x	t
	2	54	15	161
	2	52	46	170
	3	84	123	16
	3	79	72	25
	3	76	51	107
	3	74	134	202
	3	71	80	212
	3	69	24	223
	4	74	104	27
	4	72	63	36
	4	75	63	109
	4	74	104	119
	4	74	106	180
	4	73	68	189
	4	81	26	199
5	1	77	113	18
	1	74	59	28
	1	71	37	124
	1	74	92	134
	1	73	97	234
	1	72	45	244
	2	86	76	28
	2	75	51	38
	2	54	18	48
	2	67	31	144
	2	67	59	152
	2	83	90	162
	3	74	91	25
	3	74	45	34
	3	75	76	119
	3	74	122	128
	3	76	106	208
	3	73	57	217
	4	73	54	29
	4	77	100	39
	4	81	138	48
	4	70	126	124
	4	71	88	133
	4	80	43	143
	4	76	56	207
	4	74	95	216
	4	83	137	226
6	1	74	115	16
	1	77	64	27
	1	77	24	36
	1	74	69	101
	1	79	115	111
	1	78	93	207
	1	80	44	218
	2	54	99	36
	2	65	48	55
	2	81	22	64
	2	82	53	143
	2	60	80	152
	2	69	103	162
	2	48	126	170
	3	71	131	14
	3	69	87	24
	3	68	46	33
	3	69	38	104
	3	69	84	114
	3	75	133	124
	3	67	79	203
	3	68	26	213
	4	75	121	22
	4	79	81	32
	4	81	38	42
	4	79	45	125
	4	79	91	135
	4	75	116	201
	4	74	76	210
	4	70	31	220
7	1	74	120	13
	1	71	60	23
	1	72	63	102
	1	75	120	112
	1	76	106	220
	1	73	54	229
	2	39	35	30
	2	50	56	38
	2	51	84	47
	2	62	118	56
	2	64	118	168
	2	52	89	177
	2	59	62	186
	2	43	34	196

Person #	Scene #	y	x	t
	3	69	127	14
	3	68	87	23
	3	68	47	32
	3	68	56	103
	3	68	97	112
	3	69	138	121
	3	70	126	201
	3	68	85	210
	3	73	44	219
	4	72	28	19
	4	74	64	29
	4	76	102	39
	4	77	138	50
	4	76	121	108
	4	76	88	116
	4	83	42	127
	4	77	35	207
	4	72	66	216
	4	76	99	226
	4	76	127	235
8	1	71	122	16
	1	68	74	26
	1	68	26	36
	1	71	32	158
	1	70	80	167
	1	73	131	177
	2	47	16	20
	2	58	34	30
	2	55	59	40
	2	67	83	50
	2	68	112	60
	2	80	141	70
	2	82	135	193
	2	69	102	203
	2	72	75	213
	2	58	49	223
	2	63	29	233
	3	71	114	17
	3	72	66	27
	3	70	33	151
	3	74	92	161
	3	71	142	170
	4	75	44	36
	4	84	86	47
	4	77	125	57
	4	84	122	130
	4	77	80	140
	4	82	39	150
	4	77	71	221
	4	86	114	231
9	1	74	112	18
	1	67	74	26
	1	66	31	35
	1	68	79	167
	1	73	136	178
	2	46	16	20
	2	60	39	32
	2	54	57	39
	2	70	89	52
	2	66	109	59
	2	84	148	72
	2	81	128	195
	2	67	102	203
	2	72	69	215
	2	57	49	223
	3	68	124	15
	3	70	71	26
	3	71	87	160
	3	69	137	169
	4	74	44	36
	4	86	90	48
	4	76	118	55
	4	85	116	131
	4	75	83	139
	4	82	34	151
	4	74	25	211
	4	75	66	220
	4	86	119	232
	4	74	148	239

Figure A.20: Jogging Coordinates Continued.

Person #	Scene #	y	x	t
1	1	91	148	10
	1	91	89	19
	1	86	33	109
	1	90	116	219
	2	103	152	11
	2	88	105	21
	2	78	59	30
	2	48	5	103
	2	66	19	110
	2	74	57	120
	2	93	99	130
	2	101	137	225
	2	88	92	235
	2	77	52	244
	3	89	124	13
	3	88	62	121
	3	104	159	219
	4	99	25	18
	4	107	91	28
	4	111	132	83
	4	105	64	93
	4	102	17	162
	4	110	82	173
10	1	89	64	201
	2	60	59	47
	2	76	23	155
	2	62	43	161
	2	54	76	168
	3	99	152	16
	3	96	53	134
	3	96	100	216
	4	96	78	33
	4	97	123	39
	4	96	83	121
	4	96	25	171
	4	97	78	178
11	1	88	118	17
	1	86	74	24
	1	95	10	155
	1	90	71	164
	2	75	158	17
	2	83	129	24
	2	92	85	32
	2	98	60	161
	2	83	108	169
	3	92	106	16
	3	92	11	152
	3	93	76	159
	4	101	128	24
	4	95	51	135
	4	96	122	220
12	1	84	76	18
	1	86	4	85
	1	81	79	93
	1	83	129	174
	2	53	151	13
	2	58	124	20
	2	80	85	30
	2	94	14	110
	2	85	53	116
	2	73	88	123
	2	60	131	132
	2	65	135	234
	2	78	95	244
	3	95	115	19
	3	84	21	112
	3	86	87	120
	3	86	137	201
	3	89	72	209
	4	92	112	27
	4	101	50	112
	4	97	124	181
	4	89	65	190
13	1	90	125	20
	1	88	78	28
	1	88	43	183
	1	96	95	192
	2	53	140	24
	2	58	102	33
	2	65	72	41
	2	75	32	170
	2	63	65	178
	2	60	111	188
	2	56	134	196
	3	98	131	19
	3	92	82	28

Person #	Scene #	y	x	t
	3	91	43	36
	3	87	43	160
	3	84	86	168
	4	98	136	30
	4	92	78	39
	4	99	36	137
	4	107	95	146
	4	98	121	226
	4	100	65	235
14	1	81	105	35
	1	84	8	168
	1	85	86	178
	2	44	142	32
	2	56	109	41
	2	70	63	50
	2	69	65	177
	2	63	115	186
	3	95	52	17
	3	85	156	173
	3	87	101	177
	3	83	71	183
	4	94	109	27
	4	95	8	129
	4	89	80	140
	4	93	106	241
15	1	88	154	17
	1	86	113	25
	1	81	60	34
	1	87	33	121
	1	90	86	131
	1	85	149	198
	1	88	99	207
	2	101	10	16
	2	84	53	24
	2	75	87	31
	2	67	123	39
	2	65	146	153
	2	80	112	162
	2	92	57	173
	3	81	147	16
	3	80	79	25
	3	80	59	124
	3	80	140	218
	3	82	79	227
	4	99	140	25
	4	96	85	34
	4	103	35	115
	4	96	102	125
	4	96	107	199
16	1	86	127	23
	1	85	58	32
	1	85	11	123
	1	85	83	132
	1	89	148	238
	1	87	93	246
	2	67	156	28
	2	71	125	36
	2	80	89	44
	2	92	14	140
	2	80	60	148
	2	70	99	156
	2	63	129	164
	3	90	154	21
	3	88	89	29
	3	84	32	152
	3	88	97	160
	4	98	123	27
	4	105	74	34
	4	98	46	130
	4	99	85	136
	4	104	141	210
	4	100	99	216
17	1	85	113	14
	1	81	40	114
	1	83	111	219
	2	73	158	14
	2	66	120	20
	2	50	80	27
	2	44	54	33
	2	40	27	40
	2	43	10	133
	2	46	32	140
	2	54	54	146
	2	56	86	153
	3	96	133	13
	3	91	65	20

Person #	Scene #	y	x	t
	3	78	24	126
	3	79	96	133
	4	94	139	25
	4	93	88	32
	4	89	66	102
	4	89	154	163
	4	90	86	172
	4	87	64	245
18	1	69	12	19
	1	71	74	29
	1	82	145	122
	1	68	91	132
	1	77	41	141
	1	77	13	233
	1	77	55	241
	2	50	20	19
	2	60	63	27
	2	88	95	35
	2	82	117	99
	2	68	73	106
	2	63	44	221
	2	77	84	229
	3	86	11	14
	3	94	56	22
	3	89	148	113
	3	80	94	121
	3	88	4	222
	3	85	63	231
	3	97	114	240
	4	99	34	24
	4	99	96	33
	4	96	101	102
	4	97	47	160
	4	100	108	170
19	1	83	129	19
	1	78	82	27
	1	76	63	119
	1	81	116	128
	1	83	158	216
	1	81	105	225
	1	80	53	235
	2	58	26	30
	2	68	45	38
	2	80	80	47
	2	117	142	143
	2	93	105	152
	2	83	70	161
	2	67	45	171
	2	61	22	180
	3	82	128	15
	3	79	71	24
	3	78	59	127
	3	78	104	135
	3	94	157	245
	4	87	153	19
	4	84	144	24
	4	84	89	33
	4	84	26	107
	4	83	83	116
	4	87	151	179
	4	84	106	187
2	1	105	149	15
	1	99	102	24
	1	103	55	33
	1	107	19	106
	1	110	53	114
	1	100	102	123
	1	106	139	180
	1	102	99	189
	1	107	56	198
	2	68	106	65
	2	78	139	75
	2	96	130	126
	2	84	96	135
	3	98	153	20
	3	101	120	31
	3	94	78	42
	3	95	48	51
	3	105	17	146
	3	100	50	155
	3	106	89	165
	3	102	119	175
	3	88	133	236
	3	92	112	244
	4	106	140	22
	4	103	76	35

Figure A.21: Running Coordinates.

Person #	Scene #	y	x	t	Person #	Scene #	y	x	t	Person #	Scene #	y	x	t
	4	111	33	47		2	41	143	32		2	80	48	136
	4	100	13	92		2	50	131	42		3	89	122	18
	4	107	52	103		2	61	110	51		3	91	52	26
	4	98	93	114		2	74	88	61		3	87	52	88
	4	106	144	190		2	88	53	71		3	88	114	164
	4	100	113	200		2	101	29	172		3	91	20	231
	4	109	71	212		2	82	63	182		4	97	84	29
20	1	78	31	13		2	69	95	191		4	92	57	87
	1	89	112	22		2	56	111	201		4	92	106	145
	1	92	118	104		3	82	125	18		4	99	43	205
	1	85	54	112		3	85	63	28		4	94	96	214
	1	89	43	215		3	85	17	124	5	1	96	151	14
	1	96	104	223		3	82	66	133		1	88	90	22
	2	37	18	19		3	83	117	248		1	91	59	100
	2	44	43	26		4	89	29	28		1	88	98	188
	2	62	69	34		4	94	83	38		2	115	151	13
	2	83	122	42		4	91	139	109		2	92	93	22
	2	79	129	102		4	94	90	119		2	80	47	30
	2	50	66	117		4	96	43	186		2	76	32	117
	2	41	21	210		4	91	103	197		2	92	84	126
	2	59	55	220	25	1	102	157	17		2	93	106	207
	3	94	73	17		1	92	106	26		2	79	51	216
	3	94	116	83		1	90	19	85		3	86	71	26
	3	93	67	90		1	94	91	95		3	89	107	113
	3	92	34	173		1	96	103	149		3	83	121	202
	3	85	92	181		1	90	29	209		4	96	66	34
	4	98	28	19		1	93	92	219		4	101	131	114
	4	105	75	26		2	61	25	28		4	101	78	121
	4	100	143	95		2	71	54	37		4	99	11	177
	4	100	82	103		2	80	74	46		4	98	58	185
	4	106	53	169		2	91	116	56	6	1	96	155	12
	4	99	94	176		2	113	134	111		1	93	99	20
	4	100	88	242		2	94	102	120		1	93	29	92
21	1	79	114	17		2	83	71	128		1	98	92	100
	1	78	31	133		2	68	49	137		1	103	144	177
	2	60	77	48		2	62	34	226		1	93	115	183
	2	66	60	54		2	69	51	235		2	64	155	13
	2	66	39	134		2	82	83	244		2	66	138	18
	2	57	71	143		3	93	129	20		2	74	106	27
	3	76	110	17		3	89	86	28		2	82	61	36
	3	82	46	133		3	83	44	86		2	98	8	99
	4	91	53	32		3	87	94	95		2	89	24	106
	4	87	111	131		3	95	137	154		2	81	58	114
	4	95	12	204		3	88	74	164		2	74	85	123
	4	90	76	214		3	85	42	223		2	65	122	132
22	1	68	157	14		4	93	51	31		2	61	154	240
	1	65	97	22		4	93	108	41		2	66	133	249
	1	76	41	146		4	91	108	114		3	88	156	13
	1	78	123	155		4	94	72	176		3	90	78	23
	1	65	133	228	3	1	79	157	12		3	88	12	82
	1	62	69	237		1	76	87	129		3	90	53	89
	2	49	107	20		1	80	114	240		3	94	129	168
	2	42	49	156		2	27	159	19		3	88	76	176
	2	56	88	165		2	33	147	27		3	94	7	237
	3	80	150	11		2	35	136	37		3	91	43	244
	3	78	74	20		2	44	115	47		4	97	152	17
	3	86	5	145		2	50	98	57		4	94	93	26
	4	91	24	29		2	58	67	66		4	97	32	91
	4	98	98	38		2	70	46	75		4	95	111	101
	4	94	103	120		2	68	30	127		4	102	155	152
	4	98	24	187		2	49	67	139		4	97	101	161
	4	100	98	196		2	49	101	149		4	100	17	229
23	1	84	134	13		2	39	114	157		4	93	81	238
	1	85	44	110		2	42	133	166	7	1	95	150	9
	1	84	120	228		3	84	143	12		1	90	121	14
	2	54	50	39		3	84	63	22		1	88	24	93
	2	74	78	45		3	71	45	109		1	88	73	99
	2	82	101	111		3	68	95	117		1	89	136	206
	2	62	69	118		3	83	132	248		2	55	14	23
	2	50	49	125		4	106	79	26		2	54	43	30
	2	57	41	214		4	110	106	98		2	59	63	36
	2	77	69	221		4	109	50	108		2	63	93	42
	3	86	156	13		4	106	35	154		2	66	151	169
	3	88	77	20		4	100	82	163		2	63	115	175
	3	87	34	86		4	97	132	232		2	60	80	182
	3	87	100	93		4	105	72	243		3	94	151	12
	3	84	89	190	4	1	84	135	14		3	93	103	20
	4	100	62	32		1	88	68	23		3	96	43	29
	4	93	121	77		1	87	19	74		3	89	7	87
	4	100	48	130		1	87	71	82		3	91	23	92
	4	95	128	184		1	89	133	158		3	87	67	100
24	1	84	93	18		1	86	73	167		3	92	112	107
	1	83	10	120		1	88	18	219		3	88	142	249
	1	85	89	130		1	88	75	228		4	95	34	21
	1	79	93	246		2	65	94	55		4	94	95	30
	2	33	158	23		2	82	61	64		4	97	128	88

Figure A.22: Running Coordinates Continued.

Person #	Scene #	y	x	t
	4	93	89	96
	4	98	28	171
	4	97	76	179
	4	94	132	245
8	1	89	125	17
	1	83	65	25
	1	86	12	112
	1	84	83	121
	1	87	130	206
	1	85	65	214
	2	58	4	16
	2	64	27	23
	2	70	60	32
	2	79	97	40
	2	92	152	125
	2	79	111	134
	2	72	70	143
	2	65	41	152
	2	57	7	210
	2	61	25	217
	2	69	52	226
	2	76	88	234
	2	88	123	242
	3	95	144	15
	3	88	83	23
	3	89	6	111
	3	92	82	120
	3	90	154	207
	3	87	88	215
	4	102	9	25
	4	101	58	35
	4	100	109	44
	4	99	145	91
	4	105	94	101
	4	101	23	152
	4	102	82	163
	4	104	128	224
	4	98	67	234
9	1	89	112	18
	1	84	66	25
	1	85	13	112
	1	84	82	121
	1	88	130	206
	1	85	66	214
	2	59	3	16
	2	64	32	24
	2	70	61	32
	2	79	97	40
	2	91	145	126
	2	80	110	134
	2	73	71	143
	2	65	40	152
	2	57	6	210
	2	62	25	217
	2	69	52	226
	2	78	93	235
	2	89	122	242
	3	97	155	14
	3	88	82	23
	3	90	5	111
	3	92	68	119
	3	90	154	207
	3	86	87	215
	4	102	7	25
	4	101	49	34
	4	101	109	44
	4	99	145	91
	4	104	106	100
	4	99	36	154
	4	102	89	164
	4	104	128	224
	4	100	73	233

Figure A.23: Running Coordinates Continued.

Person #	Scene #	y	x	t
1	1	99	128	26
	1	97	90	40
	1	96	51	55
	1	102	19	176
	1	102	60	190
	1	105	98	204
	2	109	111	42
	2	102	79	56
	2	91	56	71
	2	86	30	86
	2	76	7	195
	2	83	23	209
	2	89	47	224
	2	95	66	237
	3	111	150	27
	3	107	114	40
	3	107	75	56
	3	97	6	191
	3	99	39	204
	3	99	75	218
	3	101	112	232
	4	105	147	36
	4	107	114	53
	4	105	79	69
	4	109	43	85
	4	107	14	185
	4	106	49	200
	4	107	89	217
	4	105	122	233
10	1	100	114	36
	1	103	78	49
	1	98	43	62
	1	103	20	240
	2	68	47	151
	2	80	36	165
	3	105	143	26
	3	102	104	39
	3	105	63	53
	3	113	19	235
	3	110	61	249
	4	105	17	27
	4	109	49	42
	4	106	86	56
	4	109	120	70
	4	103	137	188
	4	105	107	201
	4	104	74	215
	4	107	39	229
11	1	99	136	22
	1	101	105	36
	1	100	75	49
	1	104	44	63
	1	105	20	170
	1	102	51	183
	1	103	79	196
	1	101	108	211
	2	63	133	67
	2	68	117	79
	2	74	106	96
	2	81	89	110
	2	89	73	124
	2	94	53	138
	2	106	36	153
	3	102	137	24
	3	100	100	37
	3	103	67	51
	3	103	27	203
	3	102	64	217
	3	103	97	230
	4	112	122	34
	4	114	84	47
	4	111	46	61
	4	114	53	182
	4	112	94	195
12	1	101	127	28
	1	99	91	42
	1	104	54	55
	1	104	43	191
	1	99	77	204
	1	101	110	219
	2	51	152	35
	2	55	138	49
	2	60	128	62
	2	63	111	76
	2	70	98	90
	2	74	79	103
	2	82	63	117

Person #	Scene #	y	x	t
	2	87	39	132
	3	101	116	32
	3	104	74	45
	3	97	46	244
	3	102	123	53
	4	105	87	68
	4	104	50	84
	4	104	35	218
	4	107	74	234
	4	103	105	248
13	1	96	140	25
	1	95	107	39
	1	97	72	53
	1	94	38	67
	1	95	31	221
	1	97	66	234
	1	97	100	249
	2	48	148	38
	2	50	131	52
	2	56	116	67
	2	59	95	81
	2	64	75	95
	2	69	52	110
	2	75	30	124
	3	89	134	31
	3	88	103	45
	3	91	72	59
	3	89	40	72
	3	90	10	230
	3	92	42	244
	4	103	136	40
	4	107	105	57
	4	104	70	70
	4	107	37	228
	4	104	74	242
14	1	95	142	47
	1	98	109	62
	1	95	76	78
	1	97	41	93
	1	92	41	243
	2	67	139	57
	2	74	118	70
	2	83	102	84
	2	92	79	99
	2	105	57	113
	3	104	121	40
	3	101	84	54
	3	103	49	68
	4	105	136	41
	4	104	99	55
	4	110	64	71
	4	107	28	192
	4	110	62	208
	4	106	98	223
	1	92	122	35
15	1	91	91	52
	1	93	63	66
	1	90	32	83
	1	94	11	243
	2	71	148	37
	2	75	132	52
	2	81	121	68
	2	85	102	82
	2	93	87	97
	2	97	64	113
	2	105	45	127
	3	103	123	26
	3	101	88	41
	3	103	53	54
	3	103	28	201
	3	101	63	216
	3	105	101	231
	4	111	136	43
	4	107	104	59
	4	108	73	76
	4	103	42	91
	4	105	35	219
	4	108	67	238
16	1	101	145	27
	1	96	110	42
	1	97	73	57
	1	39	95	71
	1	92	35	199
	1	91	72	213
	1	93	104	228
	2	62	101	86
	2	71	86	100

Person #	Scene #	y	x	t
	2	77	62	114
	2	87	40	129
	3	100	140	34
	3	103	102	48
	3	100	61	63
	4	104	114	54
	4	106	74	69
	4	105	35	202
	4	107	74	220
	4	104	111	236
17	1	99	127	45
	1	98	64	73
	1	95	34	88
	1	91	15	190
	1	94	45	204
	1	93	74	219
	1	97	106	232
	2	50	11	31
	2	53	26	46
	2	58	38	63
	2	61	53	75
	2	65	67	90
	2	75	106	118
	2	79	129	132
	3	99	144	31
	3	100	115	46
	3	98	49	74
	3	95	39	243
	4	102	113	48
	4	106	78	63
	4	104	41	76
	4	102	28	183
	4	106	62	198
	4	104	97	213
18	1	96	44	41
	1	92	78	55
	1	96	111	68
	1	95	140	145
	1	88	108	158
	1	86	77	174
	1	80	47	187
	2	74	40	52
	2	84	61	65
	2	94	89	79
	2	108	113	92
	2	109	121	194
	2	100	86	208
	2	84	66	221
	2	78	39	235
	3	90	119	33
	3	85	87	47
	3	86	57	60
	3	92	26	193
	3	90	56	207
	3	96	87	220
	3	96	119	233
	4	114	51	24
	4	111	89	38
	4	115	126	53
	4	113	134	151
	4	110	95	164
	4	113	59	180
19	1	94	132	22
	1	97	100	36
	1	93	68	49
	1	94	37	63
	1	94	33	154
	1	96	62	168
	1	94	91	180
	1	97	122	194
	2	80	21	37
	2	88	37	49
	2	105	72	75
	2	107	105	244
	3	100	18	22
	3	99	52	34
	3	102	85	46
	3	100	120	58
	3	101	145	158
	3	100	110	171
	3	96	76	183
	3	98	42	197
	4	133	101	44
	4	100	105	58
	4	105	82	73
	4	104	51	87
	4	110	36	226

Figure A.24: Walking Coordinates.

Person #	Scene #	y	x	t
	4	110	70	240
2	1	104	141	26
	1	108	102	39
	1	103	61	52
	1	111	10	170
	1	103	49	182
	1	105	88	193
	1	103	127	206
	2	58	77	128
	2	65	93	142
	2	75	104	156
	2	84	122	173
	2	95	132	189
	3	104	115	39
	3	97	82	52
	3	99	43	67
	3	103	14	195
	3	99	51	208
	3	105	85	223
	3	103	119	236
	4	105	138	38
	4	102	101	52
	4	107	68	68
	4	102	31	83
	4	103	49	153
	4	101	88	169
	4	103	125	185
20	1	99	30	26
	1	100	63	39
	1	95	96	54
	1	96	130	187
	1	98	100	202
	1	96	68	216
	1	98	37	232
	2	48	21	60
	2	53	34	73
	2	60	42	87
	2	66	58	102
	2	74	71	114
	2	83	92	128
	2	95	111	141
	3	98	123	31
	3	95	84	43
	3	98	47	57
	3	99	27	132
	3	96	68	145
	3	99	106	159
	4	103	45	49
	4	106	83	66
	4	103	118	80
	4	104	114	188
	4	109	79	204
	4	106	40	217
21	1	102	142	26
	1	99	101	41
	1	103	59	56
	1	101	23	212
	1	98	63	227
	1	100	102	242
	2	79	77	85
	2	90	48	100
	2	94	34	220
	2	81	63	236
	3	106	116	29
	3	103	69	45
	3	107	43	220
	3	112	90	235
	4	110	26	61
	4	108	67	78
	4	110	108	96
	4	111	119	217
	4	108	76	234
22	1	99	135	34
	1	93	104	49
	1	96	68	65
	1	90	38	238
	2	36	31	52
	2	44	42	68
	2	50	60	82
	2	59	76	97
	2	64	103	113
	2	76	125	130
	3	92	126	37
	3	89	91	51
	3	93	55	67
	4	106	19	35
	4	105	57	51

Person #	Scene #	y	x	t
	4	108	94	67
	4	107	101	181
	4	111	67	197
23	1	104	127	9
	1	103	92	24
	1	106	61	40
	1	106	15	182
	1	100	47	197
	1	103	78	213
	1	102	109	227
	2	46	9	40
	2	57	23	65
	2	63	33	80
	2	71	40	92
	2	78	55	106
	2	89	66	120
	2	101	84	133
	3	101	109	35
	3	98	70	48
	3	98	18	154
	3	102	55	168
	3	100	95	182
	4	104	43	37
	4	102	78	52
	4	100	145	158
	4	104	117	172
	4	102	83	186
	4	104	50	200
24	1	99	114	32
	1	104	75	46
	1	103	14	192
	1	102	53	205
	1	106	91	218
	2	70	102	114
	2	78	84	128
	2	88	72	140
	2	98	48	152
	3	98	113	35
	3	97	77	48
	3	101	39	61
	3	95	45	199
	3	99	80	212
	3	98	116	225
	4	103	52	38
	4	102	87	51
	4	104	125	65
	4	103	123	134
	4	101	89	146
	4	103	24	233
	4	101	57	247
25	1	108	134	29
	1	103	102	44
	1	104	65	59
	1	101	27	187
	1	105	59	202
	1	103	94	217
	1	107	127	233
	2	60	41	69
	2	67	58	83
	2	75	70	96
	2	82	90	110
	2	90	108	124
	2	98	134	138
	2	111	140	220
	2	103	121	233
	2	95	97	247
	3	101	128	39
	3	101	92	54
	3	98	60	68
	3	97	26	164
	3	96	57	180
	3	99	85	194
	3	98	119	209
	4	103	37	42
	4	98	72	57
	4	102	109	72
	4	103	129	218
	4	101	97	232
	4	103	63	249
3	1	34	127	37
	1	93	98	51
	1	94	67	67
	1	87	41	80
	2	34	155	36
	2	37	147	48
	2	41	140	62
	2	44	130	76

Person #	Scene #	y	x	t
	2	44	123	91
	2	56	114	106
	2	60	103	120
	2	65	89	134
	2	71	80	148
	2	73	62	163
	2	83	48	177
	3	82	140	33
	3	80	118	47
	3	78	97	63
	3	83	75	76
	3	82	52	91
	3	81	32	105
	4	103	31	31
	4	104	73	46
	4	102	108	62
	4	104	147	144
	4	109	114	162
	4	107	80	176
	4	108	42	193
4	1	98	141	25
	1	97	106	38
	1	98	67	52
	1	94	11	135
	1	96	47	148
	1	95	83	162
	2	79	58	120
	2	94	34	134
	3	101	119	29
	3	102	75	43
	3	105	14	154
	3	106	54	167
	3	102	98	181
	4	108	139	37
	4	107	103	51
	4	109	63	66
	4	106	23	142
	4	107	63	156
	4	108	100	171
	4	113	143	246
5	1	106	140	25
	1	108	103	37
	1	102	63	51
	1	101	23	164
	1	105	61	178
	1	100	100	191
	2	109	139	26
	2	101	104	40
	2	87	85	53
	2	82	59	65
	2	75	44	79
	2	69	24	240
	3	94	120	34
	3	96	86	48
	3	94	50	60
	3	101	29	196
	3	99	65	210
	3	103	101	223
	4	100	21	36
	4	105	54	50
	4	102	91	64
	4	103	125	79
	4	99	129	171
	4	104	100	184
	4	100	66	198
6	1	107	148	21
	1	108	106	33
	1	105	65	47
	1	109	38	148
	1	110	82	162
	2	43	151	29
	2	47	143	41
	2	50	125	58
	2	56	116	71
	2	60	93	86
	2	65	83	99
	2	68	64	114
	2	74	55	127
	3	102	107	31
	3	100	66	44
	3	103	22	123
	3	103	69	136
	3	106	111	148
	4	101	144	32
	4	107	109	46
	4	102	67	61
	4	105	36	165

Figure A.25: Walking Coordinates Continued.

Person #	Scene #	y	x	t
	4	111	77	180
	4	105	117	193
7	1	108	113	27
	1	110	71	39
	2	63	11	25
	2	69	34	37
	2	74	53	48
	2	79	80	59
	2	85	102	71
	2	90	129	82
	2	88	142	205
	2	84	119	217
	2	82	89	229
	2	78	70	239
	3	91	137	24
	3	92	103	38
	3	93	71	51
	3	93	39	63
	3	97	14	132
	3	96	49	146
	3	94	82	158
	3	94	114	172
	4	103	123	36
	4	102	92	50
	4	104	58	65
	4	99	20	186
	4	101	50	201
	4	100	83	215
	4	103	113	230
8	1	93	129	23
	1	94	90	36
	1	94	50	49
	1	97	22	231
	1	94	65	244
	2	51	16	27
	2	58	34	40
	2	61	54	52
	2	66	68	63
	2	73	95	76
	2	83	117	88
	3	102	134	22
	3	100	88	35
	3	94	21	216
	3	95	66	228
	3	100	111	241
	4	109	44	35
	4	107	84	50
	4	107	124	66
	4	108	128	146
	4	107	84	161
	4	107	43	174
9	1	93	129	23
	1	94	90	36
	1	94	49	49
	1	97	21	231
	1	96	60	243
	2	52	17	28
	2	57	33	40
	2	63	51	51
	2	66	68	63
	2	73	94	76
	2	83	117	88
	3	102	134	22
	3	100	88	35
	3	94	19	216
	3	95	65	228
	3	100	111	241
	4	109	44	35
	4	107	84	50
	4	108	123	66
	4	108	128	146
	4	107	84	160
	4	107	43	174

Figure A.26: Walking Coordinates Continued.